

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶: A61K 38/18, 39/395, 38/17, 35/12 // C07K 14/475, 16/22, C12N 15/12, C12P 21/08	A2	(11) International Publication Number: WO 99/39729 (43) International Publication Date: 12 August 1999 (12.08.99)
(21) International Application Number: PCT/US99/02390 (22) International Filing Date: 3 February 1999 (03.02.99) (30) Priority Data: 09/020,598 4 February 1998 (04.02.98) US (71) Applicants: GENENTECH, INC. [US/US]; 1 DNA Way, South San Francisco, CA 94080-4990 (US). UNIVERSITY OF IOWA RESEARCH FOUNDATION [US/US]; Technology Innovation Center, Room 214 TIC, 100 Oakdale Campus, Iowa City, IA 52242-5000 (US). (72) Inventors: SLIWKOWSKI, Mark; 42 Oak Creek Lane, Iowa City, IA 94070 (US). KERN, Jeffrey, A.; 4344 Maier Avenue, Iowa City, IA 52240 (US). (74) Agents: SCHWARTZ, Timothy, R. et al.; Genentech, Inc., 1 DNA Way, South San Francisco, CA 94080-4990 (US).		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG). Published <i>Without international search report and to be republished upon receipt of that report.</i>
(54) Title: USE OF HEREGULIN AS AN EPITHELIAL CELL GROWTH FACTOR (57) Abstract Ligands which bind to the HER2, HER3 and/or HER4 receptors are useful as normal epithelial cell growth factors.		

USE OF HEREGULIN AS AN EPITHELIAL CELL GROWTH FACTOR

BACKGROUND OF THE INVENTION**Field of the Invention**

5 This invention relates to the use of HER2, HER3 and/or HER4 ligands, in particular heregulin polypeptides, as epithelial cell growth factors.

Description of Background and Related Art

10 The HER (ErbB) family belongs to the subclass I receptor tyrosine kinase superfamily and consists of three distinct receptors, HER2, HER3, and HER4. A ligand for this ErbB family is the protein heregulin (HRG), a multidomain containing protein with at least 15 distinct isoforms.

15 Transduction of signals that regulate cell growth and differentiation is regulated in part by phosphorylation of various cellular proteins. Protein tyrosine kinases are enzymes that catalyze this process. Receptor protein tyrosine kinases are believed to direct cellular growth via ligand-stimulated tyrosine phosphorylation of intracellular substrates. Growth factor receptor protein tyrosine kinases of the class I subfamily include the 170 kDa epidermal growth factor receptor (EGFR) encoded by the *erbB1* gene. *erbB1* has been causally implicated in human malignancy. In particular, increased expression of this gene has been observed in more aggressive carcinomas of the breast, bladder, lung and stomach.

20 The second member of the class I subfamily, p185^{neu} was originally identified as the product of the transforming gene from neuroblastomas of chemically treated rats. The *neu* gene (also called *erbB2* and HER2) encodes a 185 kDa receptor protein tyrosine kinase. Amplification and/or overexpression of the human HER2 gene correlates with a poor prognosis in breast and ovarian cancers (Slamon *et al.*, *Science* 235:177-182 (1987); and Slamon *et al.*, *Science* 244:707-712 (1989)). Overexpression of HER2 has been correlated with other carcinomas including carcinomas of the stomach, endometrium, salivary gland, lung, 25 kidney, colon and bladder. Accordingly, Slamon *et al.* in US Pat No. 4,968,603 describe and claim various diagnostic assays for determining HER2 gene amplification or expression in tumor cells. Slamon *et al.* discovered that the presence of multiple gene copies of HER2 oncogene in tumor cells indicates that the disease is more likely to spread beyond the primary tumor site, and that the disease may therefore require more aggressive treatment than might otherwise be indicated by other diagnostic factors. Slamon *et al.* conclude that the HER2 gene amplification test, together with the determination of lymph node status, 30 provides greatly improved prognostic utility.

A further related gene, called *erbB3* or HER3, has also been described. See US Pat. No. 5,183,884; Kraus *et al.*, *Proc. Natl. Acad. Sci. USA* 86:9193-9197 (1989); EP Pat Appln No 444,961A1; and Kraus *et al.*, *Proc. Natl. Acad. Sci. USA* 90:2900-2904 (1993). Kraus *et al.* (1989) discovered that markedly elevated 35 levels of *erbB3* mRNA were present in certain human mammary tumor cell lines indicating that *erbB3*, like *erbB1* and *erbB2*, may play a role in human malignancies. Also, Kraus *et al.* (1993) showed that EGF-dependent activation of the ErbB3 catalytic domain of a chimeric EGFR/ErbB3 receptor resulted in a proliferative response in transfected NIH-3T3 cells. This is now believed to be the result of endogenous ErbB1 or ErbB2 in NIH-3T3. Furthermore, these researchers demonstrated that some human mammary tumor cell lines display a significant elevation of steady-state ErbB3 tyrosine phosphorylation further 40 indicating that this receptor may play a role in human malignancies. The role of *erbB3* in cancer has been explored by others. It has been found to be overexpressed in breast (Lemoine *et al.*, *Br. J. Cancer* 66:1116-1121 (1992)), gastrointestinal (Poller *et al.*, *J. Pathol.* 168:275-280 (1992), Rajkumar *et al.*, *J. Pathol.* 170:271-278 (1993), and Sanidas *et al.*, *Int. J. Cancer* 54:935-940 (1993)), and pancreatic cancers (Lemoine *et al.*, *J. Pathol.* 168:269-273 (1992), and Friess *et al.*, *Clinical Cancer Research* 1:1413-1420 (1995)). 45

The class I subfamily of growth factor receptor protein tyrosine kinases has been further extended to

do not have the complete glycosylated spacer region between the Ig-like domain and EGF-like domain. Only one of the GGFs, GGFII, possessed a N-terminal signal peptide. See also WO 92/18627; WO 94/00140; WO 94/04560; WO 94/26298; and WO 95/32724 which refer to GGFs and uses thereof.

Ho *et al.* in *J. Biol. Chem.* 270(4):14523-14532 (1995) describe another member of the heregulin family called sensory and motor neuron-derived factor (SMDF). This protein has an EGF-like domain characteristic of all other heregulin polypeptides but a distinct N-terminal domain. The major structural difference between SMDF and the other heregulin polypeptides is the lack in SMDF of the Ig-like domain and the "glyco" spacer characteristic of all the other heregulin polypeptides. Another feature of SMDF is the presence of two stretches of hydrophobic amino acids near the N-terminus.

While the heregulin polypeptides were first identified based on their ability to activate the HER2 receptor (see Holmes *et al.*, *supra*), it was discovered that certain ovarian cells expressing *neu* and *neu*-transfected fibroblasts did not bind or crosslink to NDF, nor did they respond to NDF to undergo tyrosine phosphorylation (Peles *et al.*, *EMBO J.* 12:961-971 (1993)). This indicated another cellular component was necessary for conferring full heregulin responsiveness. Carraway *et al.* subsequently demonstrated that ¹²⁵I-

rHRGβ1₁₇₇₋₂₄₄ bound to NIH-3T3 fibroblasts stably transfected with bovine *erbB3* but not to non-transfected parental cells. Accordingly, they conclude that ErbB3 is a receptor for HRG and mediates phosphorylation of intrinsic tyrosine residues as well as phosphorylation of ErbB2 receptor in cells which express both receptors. Carraway *et al.*, *J. Biol. Chem.* 269(19):14303-14306 (1994). Sliwkowski *et al.*, *J. Biol. Chem.* 269(20):14661-14665 (1994) found that cells transfected with HER3 alone show low affinities for heregulin, whereas cells transfected with both HER2 and HER3 show higher affinities.

This observation correlates with the "receptor cross-talking" described previously by Kokai *et al.*, *Cell* 58:287-292 (1989); Stern *et al.*, *EMBO J.* 7:995-1001 (1988); and King *et al.*, 4:13-18 (1989). These researchers found that binding of EGF to the EGFR resulted in activation of the EGFR kinase domain and cross-phosphorylation of p185^{HER2}. This is believed to be a result of ligand-induced receptor heterodimerization and the concomitant cross-phosphorylation of the receptors within the heterodimer (Wada *et al.*, *Cell* 61:1339-1347 (1990)).

Plowman and his colleagues have similarly studied p185^{HER4}/p185^{HER2} activation. They expressed p185^{HER2} alone, p185^{HER4} alone, or the two receptors together in human T lymphocytes and demonstrated that heregulin is capable of stimulating tyrosine phosphorylation of p185^{HER4}, but could only stimulate p185^{HER2} phosphorylation in cells expressing both receptors. Plowman *et al.*, *Nature* 336:473-475 (1993).

The biological role of heregulin has been investigated by several groups. For example, Falls *et al.*, (discussed above) found that ARIA plays a role in myotube differentiation, namely affecting the synthesis and concentration of neurotransmitter receptors in the postsynaptic muscle cells of motor neurons. Corfas and Fischbach demonstrated that ARIA also increases the number of sodium channels in chick muscle. Corfas and Fischbach, *J. Neuroscience*, 13(5): 2118-2125 (1993). It has also been shown that GGFII is mitogenic for subconfluent quiescent human myoblasts and that differentiation of clonal human myoblasts in the continuous presence of GGFII results in greater numbers of myotubes after six days of differentiation (Sklar *et al.*, *J. Cell Biochem.*, Abst. W462, 18D, 540 (1994)). See also WO 94/26298 published November 24, 1994.

Holmes *et al.*, *supra*, found that HRG exerted a mitogenic effect on mammary cell lines (such as SK-BR-3 and MCF-7). The mitogenic activity of GGFs on Schwann cells has also been reported. See, e.g., Brockes *et al.*, *J. Biol. Chem.* 255(18):8374-8377 (1980); Lemke and Brockes, *J. Neurosci.* 4:75-83 (1984); Lemke and Brockes, *J. Neurosci.* 4:75-83 (1984); Brockes *et al.*, *Ann. Neurol.* 20(3):317-322 (1986); Brockes, J., *Methods in Enzym.*, 147: 217-225 (1987) and Marchionni *et al.*, *supra*. Schwann cells constitute

87:6378-6382 (1990). However, early efforts to select high affinity fusion phage failed, presumably due to the polyvalence of the phage particles. This problem was solved with the development of a "monovalent" phage display system in which the fusion protein is expressed at a low level from a phagemid and a helper phage provides a large excess of wild-type coat protein. Bass *et al.*, *Proteins* 8:309-314 (1990); Lowman *et al.*, *Biochem.* 30:10832-10838 (1991). Monovalent phage display can be used to generate and screen a large number of variant polypeptides to isolate those that bind with high affinity to a target of interest.

Approximately 50,000 infants are born in the United States every year with birth weights, less than 1.5 kg. About two thirds of these very low birth weight infants have evidence of pulmonary immaturity manifested as respiratory distress shortly after birth. The majority of these infants require mechanical ventilation. Respiratory distress syndrome, caused by insufficient pulmonary surfactant production, as well as structural immaturity of the lung, is responsible for respiratory difficulties observed in these prematurely born neonates. Well developed alveoli are necessary to provide efficient oxygen transfer from the air-liquid interface of the lung to the systemic circulation. Surfactant proteins are critical in reducing the alveolar surface tension at low lung volumes and preventing alveolar collapse.

A need continues to exist for a method of treatment for respiratory distress syndrome and other diseases associated with immature lung development and low lung surfactant production.

SUMMARY OF THE INVENTION

In general an object of the invention is to provide a method of inducing epithelial cell growth and development for the purpose of promoting repair and healing of tissue damage or injury.

Accordingly, one object of this invention is to provide a method of treating respiratory distress syndrome in patients, primarily human patients, in need of such treatment. A further object is to provide a method of inducing lung epithelial cell growth and development. A further object is to provide a method of increasing lung surfactant protein A production in the lung of persons with impaired oxygen transfer in the lung alveoli. This invention is useful in treating infants/neonates with respiratory distress as well as youth and adults with poor lung function due to lung injury or damage.

In one aspect of this invention, it has now been discovered that these objects and the broader objective of treating conditions associated with epithelial cell damage and injury are achieved by administering to a patient in need of such treatment an effective amount of a heregulin ligand, preferably a polypeptide or fragment thereof. These heregulin (HRG) polypeptides, include HRG- α , HRG- β 1, HRG- β 2, HRG- β 3 and other HRG polypeptides which cross-react with antibodies directed against these family members and/or which are substantially homologous as defined below and includes HRG variants such as N-terminal and C-terminal fragments thereof. A preferred HRG is the ligand disclosed in Fig. 1A - 1D and further designated HRG- α . Other preferred HRGs are the ligands disclosed in Figure 2A - 2E, and designated HRG- β 1; disclosed in Figure 3A - 3E designated HRG- β 2; and disclosed in Figure 4A - 4C designated HRG- β 3.

In another aspect, the invention provides a method in which HRG agonist antibodies are administered to achieve the objects of the invention. In this embodiment, HER2/HER3 or fragments thereof (which also may be synthesized by *in vitro* methods) are fused (by recombinant expression or an *in vitro* peptidyl bond) to an immunogenic polypeptide and this fusion polypeptide, in turn, is used to raise antibodies against a HER2/HER3 epitope. Agonist antibodies are recovered from the serum of immunized animals. Alternatively, monoclonal antibodies are prepared from *in vitro* cells or *in vivo* immunized animals in conventional fashion. If desired, the agonist antibodies may be obtained by phage display selection from a phage library of antibodies or antibody fragments. Preferred antibodies identified by routine screening will bind to the receptor, but will not substantially cross-react with any other known ligands such as EGF, and will activate the HER receptors HER2, HER3 and/or HER4. In addition, antibodies may be selected that are

ID NO:12) of γ -HRG obtained as described in U.S. Serial No. 08/891,845. The hydrophobic region is underlined. The EGF-like domain is shaded, cysteine residues in the EGF-like domain are circled. N-linked glycosylation sites are marked above the nucleic acid sequence with a ().

Figure 8 shows the cDNA sequence (SEQ ID NO:13) and amino acid sequence (SEQ ID NO:14) of SMDF obtained as described in U.S. Serial No. 08/339,517. An EGF-like domain and the apolar and uncharged domains (*i.e.* "apolar I" consisting of residues from about 48-62 and "apolar II" consisting of residues from about 76-100) are underlined. Cysteines in the EGF-like domain and in the "cysteine knot" in the unique N-terminal domain ("NTD-cys knot") are boxed. The stop codon is denoted by the letter "O".

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

HRG ligands, in particular polypeptides and agonist antibodies thereof, have affinity for and stimulate the HER2, HER3 and/or HER4 receptors or combinations thereof in autophosphorylation. Included within the definition of HRG ligands, in addition to HRG- α , HRG- β 1, HRG- β 2, HRG- β 3 and HRG- β 2-like, are other polypeptides binding to the HER2, HER3 and/or HER4 receptor, which bear substantial amino acid sequence homology to HRG- α or HRG- β 1. Such additional polypeptides fall within the definition of HRG as a family of polypeptide ligands that bind to the HER2, HER3 and/or HER4 receptors.

Heregulin polypeptides bind with varying affinities to the HER2, HER3 and/or HER4 receptors. Generally, the HER3 and HER4 receptors are bound with high affinity. It is also known that heterodimerization of HER2 with HER3 and of HER2 with HER4 occurs with subsequent receptor cross-phosphorylation as described above. In the present invention, epithelial cell growth and/or proliferation is induced when a heregulin protein interacts and binds with an individual receptor molecule or a receptor dimer such that receptor phosphorylation is induced. Binding and activation of HER2, HER3, HER4 or combinations thereof, therefore, is meant to include activation of any form of the receptor necessary for receptor activation and biologic function including monomeric receptor and dimeric receptor forms. Dimeric receptor forms may be referred to below, for example, as HER2/HER3, HER2/HER4, and HER3/HER4.

I. DEFINITIONS

In general, the following words or phrases have the indicated definition when used in the description, examples, and claims.

"Heregulin" (HRG) ligand is defined herein to be any isolated ligand, preferably a polypeptide sequence which possesses a biological property of a naturally occurring HRG polypeptide. Ligands within the scope of this invention include the NDF, ARIA and GGF growth factor heregulin proteins identified above as well as the SMDF and HRG polypeptides discussed in detail herein. These isolated NDF, ARIA and GGF heregulin polypeptides are well known in the art. HRG includes the polypeptides shown in Figs. 1A-1D, 2A-2E, 3A-3E, 4A-4C, 5A-5D, 6A-6C, 7A-7C and 8 and mammalian analogues thereof. Included are HRG variants such as the γ -HRG described in WO 98/02541, published 22 January 1998; the variants described in WO 98/35036, published 13 August 1998; and the SMDF variants described in US Patent NO. 5,770,567 granted 23 June 1998 (WO 96/15244). These applications are incorporated herein in their entirety. These variants can be prepared by the methods described below, optionally together with alanine scanning and phage display techniques known in the art. Cunningham and Wells, Science 244:1081-85 (1989); Bass et al., Proteins 8:309-14 (1990); Lowman et al., Biochem. 30:10832-38 (1991).

The term a "normal" epithelial cell means an epithelial cell which is not transformed, *i.e.*, is non-cancerous and/or non-immortalized. Further, the normal epithelial cell is preferably not aneuploid. Aneuploidy exists when the nucleus of a cell does not contain an exact multiple of the haploid number of chromosomes, one or more chromosomes being present in greater or lesser number than the rest. Typical properties of transformed cells which fall outside the scope of this invention include the ability to form

in its native conformation generally is HRG as found in nature which has not been denatured by chaotropic agents, heat or other treatment that substantially modifies the three dimensional structure of HRG as determined, for example, by migration on nonreducing, nondenaturing sizing gels. Antibody used in this determination may be rabbit polyclonal antibody raised by formulating native HRG from a non-rabbit species in Freund's complete adjuvant, subcutaneously injecting the formulation, and boosting the immune response by intraperitoneal injection of the formulation until the titer of anti-HRG antibody plateaus.

Ordinarily, biologically or antigenically active HRG will have an amino acid sequence having at least 75% amino acid sequence identity with a given HRG sequence, more preferably at least 80%, even more preferably at least 90%, and most preferably at least 95%. Identity or homology with respect to an HRG sequence is defined herein as the percentage of amino acid residues in the candidate sequence that are identical with HRG residues in the HRG of Fig. 6A-6C, after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent homology, and not considering any conservative substitutions as part of the sequence identity. None of N-terminal, C-terminal or internal extensions, deletions, or insertions into HRG sequence shall be construed as affecting homology.

Thus, the biologically active and antigenically active HRG polypeptides that are the subject of this invention include each entire HRG sequence; fragments thereof having a consecutive sequence of at least 5, 10, 15, 20, 25, 30 or 40 amino acid residues from HRG sequence; amino acid sequence variants of HRG sequence wherein an amino acid residue has been inserted N- or C-terminal to, or within, HRG sequence or its fragment as defined above; amino acid sequence variants of HRG sequence or its fragment as defined above has been substituted by another residue. HRG polypeptides include those containing predetermined mutations by, e.g., site-directed or PCR mutagenesis, and other animal species of HRG polypeptides such as rabbit, rat, porcine, non-human primate, equine, murine, and ovine HRG and alleles or other naturally occurring variants of the foregoing and human sequences; derivatives of HRG or its fragments as defined above wherein HRG or its fragments have been covalently modified by substitution, chemical, enzymatic, or other appropriate means with a moiety other than a naturally occurring amino acid (for example a detectable moiety such as an enzyme or radioisotope); glycosylation variants of HRG (insertion of a glycosylation site or deletion of any glycosylation site by deletion, insertion or substitution of appropriate amino acid); and soluble forms of HRG, such as HRG-GFD or those that lack a functional transmembrane domain.

"Isolated" means a ligand, such as HRG, which has been identified and separated and/or recovered from a component of its natural environment. Contaminant components of its natural environment are materials which would interfere with diagnostic or therapeutic uses for HRG, and may include proteins, hormones, and other substances. In preferred embodiments, HRG will be purified (1) to greater than 95% by weight of protein as determined by the Lowry method or other validated protein determination method, and most preferably more than 99% by weight, (2) to a degree sufficient to obtain at least 15 residues of N-terminal or internal amino acid sequence by use of the best commercially available amino acid sequenator marketed on the filing date hereof, or (3) to homogeneity by SDS-PAGE using Coomassie blue or, preferably, silver stain. Isolated HRG includes HRG in situ within recombinant cells since at least one component of HRG natural environment will not be present. Isolated HRG includes HRG from one species in a recombinant cell culture of another species since HRG in such circumstances will be devoid of source polypeptides. Ordinarily, however, isolated HRG will be prepared by at least one purification step.

In accordance with this invention, HRG nucleic acid is RNA or DNA containing greater than ten bases that encodes a biologically or antigenically active HRG, is complementary to nucleic acid sequence encoding such HRG, or hybridizes to nucleic acid sequence encoding such HRG and remains stably bound to it under stringent conditions.

Preferably, HRG nucleic acid encodes a polypeptide sharing at least 75% sequence identity, more preferably at least 80%, still more preferably at least 85%, even more preferably at 90%, and most preferably

"Restriction enzyme digestion" of DNA refers to catalytic cleavage of the DNA with an enzyme that acts only at certain locations in the DNA. Such enzymes are called restriction endonucleases, and the sites for which each is specific is called a restriction site. The various restriction enzymes used herein are commercially available and their reaction conditions, cofactors, and other requirements as established by the enzyme suppliers are used. Restriction enzymes commonly are designated by abbreviations composed of a capital letter followed by other letters representing the microorganism from which each restriction enzyme originally was obtained, and then a number designating the particular enzyme. In general, about 1 mg of plasmid or DNA fragment is used with about 1-2 units of enzyme in about 20 ml of buffer solution. Appropriate buffers and substrate amounts for particular restriction enzymes are specified by the manufacturer. Incubation of about 1 hour at 37°C is ordinarily used, but may vary in accordance with the supplier's instructions. After incubation, protein or polypeptide is removed by extraction with phenol and chloroform, and the digested nucleic acid is recovered from the aqueous fraction by precipitation with ethanol. Digestion with a restriction enzyme may be followed with bacterial alkaline phosphatase hydrolysis of the terminal 5' phosphates to prevent the two restriction cleaved ends of a DNA fragment from "circularizing" or forming a closed loop that would impede insertion of another DNA fragment at the restriction site. Unless otherwise stated, digestion of plasmids is not followed by 5' terminal dephosphorylation. Procedures and reagents for dephosphorylation are conventional as described in sections 1.56-1.61 of Sambrook *et al.*, (*Molecular Cloning: A Laboratory Manual* New York: Cold Spring Harbor Laboratory Press, 1989).

"Recovery" or "isolation" of a given fragment of DNA from a restriction digest means separation of the digest on polyacrylamide or agarose gel by electrophoresis, identification of the fragment of interest by comparison of its mobility versus that of marker DNA fragments of known molecular weight, removal of the gel section containing the desired fragment, and separation of the gel from DNA. This procedure is known generally. For example, see Lawn *et al.*, *Nucleic Acids Res.* 9:6103-6114 (1981), and Goeddel *et al.*, *Nucleic Acids Res.* 8:4057 (1980).

"Northern analysis" is a method used to identify RNA sequences that hybridize to a known probe such as an oligonucleotide, DNA fragment, cDNA or fragment thereof, or RNA fragment. The probe is labeled with a radioisotope such as ³²P, or by biotinylation, or with an enzyme. The RNA to be analyzed is usually electrophoretically separated on an agarose or polyacrylamide gel, transferred to nitrocellulose, nylon, or other suitable membrane, and hybridized with the probe, using standard techniques well known in the art such as those described in sections 7.39-7.52 of Sambrook *et al.*, *supra*.

"Ligation" refers to the process of forming phosphodiester bonds between two nucleic acid fragments. To ligate the DNA fragments together, the ends of the DNA fragments must be compatible with each other. In some cases, the ends will be directly compatible after endonuclease digestion. However, it may be necessary to first convert the staggered ends commonly produced after endonuclease digestion to blunt ends to make them compatible for ligation. To blunt the ends, the DNA is treated in a suitable buffer for at least 15 minutes at 15°C with about 10 units of the Klenow fragment of DNA polymerase I or T4 DNA polymerase in the presence of the four deoxyribonucleotide triphosphates. The DNA is then purified by phenol-chloroform extraction and ethanol precipitation. The DNA fragments that are to be ligated together are put in solution in about equimolar amounts. The solution will also contain ATP, ligase buffer, and a ligase such as T4 DNA ligase at about 10 units per 0.5 mg of DNA. If the DNA is to be ligated into a vector, the vector is first linearized by digestion with the appropriate restriction endonuclease(s). The linearized fragment is then treated with bacterial alkaline phosphatase, or calf intestinal phosphatase to prevent self-ligation during the ligation step.

"Preparation" of DNA from cells means isolating the plasmid DNA from a culture of the host cells. Commonly used methods for DNA preparation are the large and small-scale plasmid preparations described

"Protein microsequencing" was accomplished based upon the following procedures. Proteins from the final HPLC step were either sequenced directly by automated Edman degradation with a model 470A Applied Biosystems gas phase sequencer equipped with a 120A PTH amino acid analyzer or sequenced after digestion with various chemicals or enzymes. PTH amino acids were integrated using the CHROMPERFECT data system (Justice Innovations, Palo Alto, CA). Sequence interpretation was performed on a VAX 11/785 Digital Equipment Corporation computer as described (Henzel *et al.*, *J. Chromatography* 404:41-52 (1987)). In some cases, aliquots of the HPLC fractions were electrophoresed on 5-20% SDS polyacrylamide gels, electrotransferred to a PVDF membrane (PROBLOTT, ABI, Foster City, CA) and stained with Coomassie Brilliant Blue (Matsudaira, P., *J. Biol. Chem.* 262:10035-10038, 1987). The specific protein was excised from the blot for N terminal sequencing. To determine internal protein sequences, HPLC fractions were dried under vacuum (SPEEDVAC), resuspended in appropriate buffers, and digested with cyanogen bromide, the lysine-specific enzyme Lys-C (Wako Chemicals, Richmond, VA) or Asp-N (Boehringer Mannheim, Indianapolis, Ind.). After digestion, the resultant peptides were sequenced as a mixture or were resolved by HPLC on a C4 column developed with a propanol gradient in 0.1% TFA before sequencing as described above.

"Antibodies" (Abs) and "immunoglobulins" (Igs) are glycoproteins having the same structural characteristics. While antibodies exhibit binding specificity to a specific antigen, immunoglobulins include both antibodies and other antibody-like molecules which lack antigen specificity. Polypeptides of the latter kind are, for example, produced at low levels by the lymph system and at increased levels by myelomas.

Papain digestion of antibodies produces two identical antigen-binding fragments, called "Fab" fragments, each with a single antigen-binding site, and a residual "Fc" fragment, whose name reflects its ability to crystallize readily. Pepsin treatment yields an $F(ab')_2$ fragment that has two antigen-combining sites and is still capable of cross-linking antigen.

"Fv" is the minimum antibody fragment which contains a complete antigen-recognition and -binding site. This region consists of a dimer of one heavy chain and one light chain variable domain in tight, non-covalent association. It is in this configuration that the three hypervariable regions of each variable domain interact to define an antigen-binding site on the surface of the V_H - V_L dimer. Collectively, the six hypervariable regions confer antigen-binding specificity to the antibody. However, even a single variable domain (or half of an Fv comprising only three hypervariable regions specific for an antigen) has the ability to recognize and bind antigen, although at a lower affinity than the entire binding site.

The Fab fragment also contains the constant domain of the light chain and the first constant domain (CH1) of the heavy chain. Fab' fragments differ from Fab fragments by the addition of a few residues at the carboxyl terminus of the heavy chain CH1 domain including one or more cysteine(s) from the antibody hinge region. Fab'-SH is the designation herein for Fab' in which the cysteine residue(s) of the constant domains bear a free thiol group. $F(ab')_2$ antibody fragments originally were produced as pairs of Fab' fragments which have hinge cysteines between them. Other chemical couplings of antibody fragments are also known.

The "light chains" of antibodies (immunoglobulins) from any vertebrate species can be assigned to one of two clearly distinct types, called kappa (κ) and lambda (λ), based on the amino acid sequences of their constant domains.

Depending on the amino acid sequence of the constant domain of their heavy chains, immunoglobulins can be assigned to different classes. There are five major classes of immunoglobulins: IgA, IgD, IgE, IgG, and IgM, and several of these may be further divided into subclasses (isotypes), e.g., IgG1, IgG2, IgG3, IgG4, IgA, and IgA2. The heavy-chain constant domains that correspond to the different classes of immunoglobulins are called α , δ , ϵ , γ , and μ , respectively. The subunit structures and three-dimensional

antigen binding. For a review of sFv see Pluckthun in *The Pharmacology of Monoclonal Antibodies*, vol. 113, Rosenberg and Moore eds. Springer-Verlag, New York, pp. 269-315 (1994).

The term "diabodies" refers to small antibody fragments with two antigen-binding sites, which fragments comprise a heavy chain variable domain (V_H) connected to a light chain variable domain (V_L) in the same polypeptide chain ($V_H - V_L$). By using a linker that is too short to allow pairing between the two domains on the same chain, the domains are forced to pair with the complementary domains of another chain and create two antigen-binding sites. Diabodies are described more fully in, for example, EP 404,097; WO 93/11161; and Hollinger *et al.*, *Proc. Natl. Acad. Sci. USA* 90:6444-6448 (1993).

The expression "linear antibodies" when used throughout this application refers to the antibodies described in Zapata *et al.* *Protein Eng.* 8(10):1057-1062 (1995). Briefly, these antibodies comprise a pair of tandem Fd segments ($V_H - C_H1 - V_H - C_H1$) which form a pair of antigen binding regions. Linear antibodies can be bispecific or monospecific.

II. USE AND PREPARATION OF HRG SEQUENCES

H. PREPARATION OF HRG SEQUENCES, INCLUDING VARIANTS

The system to be employed in preparing HRG sequence will depend upon the particular HRG sequence selected. If the sequence is sufficiently small HRG may be prepared by *in vitro* polypeptide synthetic methods. Most commonly, however, HRG will be prepared in recombinant cell culture using the host-vector systems described below. Suitable HRG includes any biologically active and antigenetically active HRG.

In general, mammalian host cells will be employed, and such hosts may or may not contain post-translational systems for processing HRG preprosequences in the normal fashion. If the host cells contain such systems then it will be possible to recover natural subdomain fragments such as HRG-GFD from the cultures. If not, then the proper processing can be accomplished by transforming the hosts with the required enzyme(s) or by supplying them in an *in vitro* method. However, it is not necessary to transform cells with the complete prepro or structural genes for a selected HRG when it is desired to only produce fragments of HRG sequences such as an HRG-GFD. For example, a start codon is ligated to the 5' end of DNA encoding an HRG-GFD, this DNA is used to transform host cells and the product expressed directly as the Met N-terminal form (if desired, the extraneous Met may be removed *in vitro* or by endogenous N-terminal demethionylases). Alternatively, HRG-GFD is expressed as a fusion with a signal sequence recognized by the host cell, which will process and secrete the mature HRG-GFD as is further described below. Amino acid sequence variants of native HRG-GFD sequences are produced in the same way.

HRG sequences located between the first N-terminal mature residue and the first N-terminal residue of HRG-GFD sequence, termed HRG-NTD, may function at least in part as an unconventional signal sequence or as a normally circulating carrier/precursor for HRG-GFD having unique biological activity.

HRG-NTD is produced in the same fashion as the full length molecule but from expression of DNA in which a stop codon is located at the C-terminus of HRG-NTD. In addition, HRG variants are expressed from DNA encoding protein in which both the GFD and NTD domains are in their proper orientation but which contain an amino acid insertion, deletion or substitution at the GFD-NTD cleavage site (located within the sequence VKC) which inhibits or prevents proteolytic cleavage of the NTD-GFD joining site *in vivo*, and wherein a stop codon is positioned at the 3' end of the GFD- encoding sequence. In an example of this group of variants (termed HRG-NTDXGFD), (1) the lysine residue found in the NTD-GFD joining sequence VKC is deleted or (preferably) substituted by another residue other than arginyl such as histidyl, alanyl, threonyl or seryl and (2) a stop codon is introduced in the sequence RCT or RCQ in place of cysteinyl, or threonyl (for HRG- α) or glutaminyl (for HRG- β).

screening genomic DNA libraries include, but are not limited to, oligonucleotides; cDNAs or fragments thereof that encode the same or a similar gene; and/or homologous genomic DNAs or fragments thereof. Screening the cDNA or genomic library with the selected probe may be conducted using standard procedures as described in chapters 10-12 of Sambrook *et al.*, *supra*.

5 An alternative means to isolate the gene encoding HRG- α is to use polymerase chain reaction (PCR) methodology as described in section 14 of Sambrook *et al.*, *supra*. This method requires the use of oligonucleotide probes that will hybridize to HRG- α . Strategies for selection of oligonucleotides are described below.

10 Another alternative method for obtaining the gene of interest is to chemically synthesize it using one of the methods described in Engels *et al.* (*Angew. Chem. Int. Ed. Engl.*, 28: 716-734, 1989). These methods include triester, phosphite, phosphoramidite and H-Phosphonate methods, PCR and other autoprimer methods, and oligonucleotide syntheses on solid supports. These methods may be used if the entire nucleic acid sequence of the gene is known, or the sequence of the nucleic acid complementary to the coding strand is available, or alternatively, if the target amino acid sequence is known, one may infer potential nucleic acid
15 sequences using known and preferred coding residues for each amino acid residue.

A preferred method of practicing this invention is to use carefully selected oligonucleotide sequences to screen cDNA libraries from various tissues, preferably human breast, colon, salivary gland, placental, fetal, brain, and carcinoma cell lines. Other biological sources of DNA encoding an heregulin-like ligand include other mammals and birds. Among the preferred mammals are members of the following
20 orders: bovine, ovine, equine, murine, and rodentia.

The oligonucleotide sequences selected as probes should be of sufficient length and sufficiently unambiguous that false positives are minimized. The actual nucleotide sequence(s) may, for example, be based on conserved or highly homologous nucleotide sequences or regions of HRG- α . The oligonucleotides may be degenerate at one or more positions. The use of degenerate oligonucleotides may be of particular
25 importance where a library is screened from a species in which preferential codon usage in that species is not known. The oligonucleotide must be labeled such that it can be detected upon hybridization to DNA in the library being screened. The preferred method of labeling is to use ^{32}P -labeled ATP with polynucleotide kinase, as is well known in the art, to radiolabel the oligonucleotide. However, other methods may be used to label the oligonucleotide, including, but not limited to, biotinylation or enzyme labeling.

30 Of particular interest is HRG- α nucleic acid that encodes a full-length polypeptide. In some preferred embodiments, the nucleic acid sequence includes the native HRG- α signal sequence. Nucleic acid having all the protein coding sequence is obtained by screening selected cDNA or genomic libraries, and, if necessary, using conventional primer extension procedures as described in section 7.79 of Sambrook *et al.*, *supra*, to detect precursors and processing intermediates of mRNA that may not have been reverse-
35 transcribed into cDNA.

HRG- α encoding DNA of Figures 1A-1D may be used to isolate DNA encoding the analogous ligand from other animal species via hybridization employing the methods discussed above. The preferred animals are mammals, particularly bovine, ovine, equine, feline, canine and rodentia, and more specifically rats, mice and rabbits.

40 **B. Amino Acid Sequence Variants of Heregulin**

Amino acid sequence variants of HRG are prepared by introducing appropriate nucleotide changes into HRG DNA, or by *in vitro* synthesis of the desired HRG polypeptide. Such variants include, for example, deletions from, or insertions or substitutions of, residues within the amino acid sequence shown for human HRG sequences. Any combination of deletion, insertion, and substitution can be made to arrive at the
45 final construct, provided that the final construct possesses the desired characteristics. Excluded from the

represent naturally occurring alleles (which will not require manipulation of HRG DNA) or predetermined mutant forms made by mutating the DNA, either to arrive at an allele or a variant not found in nature. In general, the location and nature of the mutation chosen will depend upon HRG characteristic to be modified. Obviously, such variations that, for example, convert HRG into a known receptor ligand, are not included within the scope of this invention, nor are any other HRG variants or polypeptide sequences that are not novel and unobvious over the prior art.

Amino acid sequence deletions generally range from about 1 to 30 residues, more preferably about 1 to 10 residues, and typically about 1 to 5 are contiguous. Deletions may be introduced into regions of low homology with other EGF family precursors to modify the activity of HRG. Deletions from HRG in areas of substantial homology with other EGF family sequences will be more likely to modify the biological activity of HRG more significantly. The number of consecutive deletions will be selected so as to preserve the tertiary structure of HRG in the affected domain, e.g., cysteine crosslinking, beta-pleated sheet or alpha helix.

Amino acid sequence insertions include amino- and/or carboxyl-terminal fusions ranging in length from one residue to polypeptides containing a hundred or more residues, as well as intrasequence insertions of single or multiple amino acid residues. Intrasequence insertions (i.e., insertions within HRG sequence) may range generally from about 1 to 10 residues, more preferably 1 to 5, and most preferably 1 to 3. Examples of terminal insertions include HRG with an N-terminal methionyl residue (an artifact of the direct expression of HRG in bacterial recombinant cell culture), and fusion of a heterologous N-terminal signal sequence to the N-terminus of HRG to facilitate the secretion of mature HRG from recombinant host cells. Such signal sequences generally will be obtained from, and thus be homologous to, the intended host cell species. Suitable sequences include STII, tPA or lpp for *E. coli*, alpha factor for yeast, and viral signals such as herpes gD for mammalian cells.

Other insertional variants of HRG include the fusion to the N- or C-terminus of HRG of an immunogenic polypeptide, e.g., bacterial polypeptides such as beta-lactamase or an enzyme encoded by the *E. coli trp* locus, or yeast protein, bovine serum albumin, and chemotactic polypeptides. C-terminal fusions of HRG-ECD with proteins having a long half-life such as immunoglobulin constant regions (or other immunoglobulin regions), albumin, or ferritin, as described in WO 89/02922, published 6 April 1989 are contemplated.

Another group of variants are amino acid substitution variants. These variants have at least one amino acid residue in HRG molecule removed and a different residue inserted in its place. The sites of greatest interest for substitutional mutagenesis include sites identified as the active site(s) of HRG, and sites where the amino acids found in HRG ligands from various species are substantially different in terms of side-chain bulk, charge, and/or hydrophobicity. A likely sub-domain of HRG-GFD having biological activity as a growth factor is the C-terminal segment, in particular within the sequence about from glycine 218 to valine 226 (HRG- α), and glycine 218 to lysine 228/serine 228 (HRG- β) based upon analogy to the EGF sub-sequence found to have EGF activity.

Other sites of interest are those in which particular residues of HRG-like ligands obtained from various species are identical. These positions may be important for the biological activity of HRG. These sites, especially those falling within a sequence of at least three other identically conserved sites, are substituted in a relatively conservative manner. Such conservative substitutions are shown in Table 1 under the heading of "preferred substitutions". If such substitutions result in a change in biological activity, then more substantial changes, denominated exemplary substitutions in Table 1, or as further described below in reference to amino acid classes, are introduced and the products screened.

In another embodiment, any methionyl residue other than the starting methionyl residue of the signal sequence, or any residue located within about three residues N- or C-terminal to each such methionyl residue, is substituted by another residue (preferably in accord with Table 1) or deleted. Alternatively, about 1-3 residues are inserted adjacent to such sites.

5 Any cysteine residues not involved in maintaining the proper conformation of HRG also may be substituted, generally with serine, to improve the oxidative stability of the molecule and prevent aberrant crosslinking.

Sites particularly suited for substitutions, deletions or insertions, or use as fragments, include, numbered from the N-terminus of HRG- α of Figure 1A - 1D:

- 10 1) potential glycosaminoglycan addition sites at the serine-glycine dipeptides at 42-43, 64-65, 151-152;
- 2) potential asparagine-linked glycosylation at positions 164, 170, 208 and 437, sites (NDS) 164-166, (NIT) 170-172, (NTS) 208-210, and NTS (609-611);
- 3) potential O-glycosylation in a cluster of serine and threonine at 209-218;
- 15 4) cysteines at 226, 234, 240, 254, 256 and 265;
- 5) transmembrane domain at 287-309;
- 6) loop 1 delineated by cysteines 226 and 240;
- 7) loop 2 delineated by cysteines 234 and 254;
- 8) loop 3 delineated by cysteines 256 and 265; and
- 20 9) potential protease processing sites at 2-3, 8-9, 23-24, 33-34, 36-37, 45-46, 48-49, 62-63, 66-67, 86-87, 110-111, 123-124, 134-135, 142-143, 272-273, 278-279 and 285-286;

Analogous regions in HRG- β 1 may be determined by reference to its' sequence. The analogous HRG- β 1 amino acids may be mutated or modified as discussed above for HRG- α . Analogous regions in HRG- β 2 may also be determined by reference to its' sequence. The analogous HRG- β 2 amino acids may be mutated or modified as discussed above for HRG- α or HRG- β 1. Analogous regions in HRG- β 3 may be determined by reference to its' sequence. Further, the analogous HRG- β 3 amino acids may be mutated or modified as discussed above for HRG- α , HRG- β 1, or HRG- β 2.

Another HRG variant is γ -HRG (or gamma-heregulin). γ -HRG is any polypeptide sequence that possesses at least one biological property of native sequence γ -HRG having SEQ ID NO:11. The biological property of this variant is the same as for HRG noted above. This variant encompasses not only the polypeptide isolated from a native γ -HRG source such as human MDA-MB-175 cells or from another source, such as another animal species, but also the polypeptide prepared by recombinant or synthetic methods. It also includes variant forms including functional derivatives, allelic variants, naturally occurring isoforms and analogues thereof. Sometimes the γ -HRG is "native γ -HRG" which refers to endogenous γ -HRG polypeptide which has been isolated from a mammal. The γ -HRG can also be "native sequence γ -HRG" insofar as it has the same amino acid sequence as a native γ -HRG (e.g. human γ -HRG shown in Fig. 7A-7C). Amino acid sequence variants of the native sequence are prepared by introducing appropriate nucleotide changes into the native sequence DNA, or by *in vitro* synthesis of the desired polypeptide. Such variants include, for example, deletions from, or insertions or substitutions of, residues within the amino acid sequence shown for the human protein in Fig. 7A-7C as generally described above for other HRG. Any combination of deletion, insertion, and substitution is made to arrive at the final construct, provided that the final construct possesses the desired characteristics. The amino acid changes also may alter post-translational processes of the native sequence, such as changing the number or position of O-linked

A183G, K185E, E186R, K187E, T188G, F197Y, M198R, D201T;
 A183G, K185E, E186R, K187E, T188G, P205Y, S206G, R207Y, Y208L, L209M;
 A183G, K185E, E186R, K187E, T188G, F197Y, M198R, D201T, P205Y, S206G, R207Y, Y208L, L209M;
 A183G, K185E, E186R, K187E, T188G, M226I;
 5 F197Y, M198R, D201T, P205Y, S206G, R207Y, Y208L, L209M;
 F197Y, M198R, D201T, P205Y, S206G, R207Y, Y208L, L209M, M226I;
 F197Y, M198R, D201T, M226I;
 A183G, K185E, E186R, K187E, T188G, F197Y, M198R, D201T, M226I;
 A183G, K185E, E186R, K187E, T188G, P205Y, S206G, R207Y, Y208L, L209M, M226I;
 10 A183G, K185E, E186R, K187E, T188G, F197Y, M198R, D201T, P205Y, S206G, R207Y, Y208L, L209M,
 M226I;
 F197Y, M198R, D201T, P205Y, S206G, R207Y, Y208L, L209M, N223H, M226I; and
 A183G, K185E, E186R, K187E, T188G, F197Y, M198R, D201T, P205Y, S206G, R207Y, Y208L, L209M,
 N223H, M226I.

15 In addition to including one or more of the amino acid substitutions disclosed herein, the heregulin
 variant can have one or more other modifications, such as an amino acid substitution, an insertion of at least
 one amino acid, a deletion of at least one amino acid, or a chemical modification. For example, the invention
 provides a heregulin variant that is a fragment. In a variation of this embodiment, the fragment includes
 residues corresponding to a portion of human heregulin- β 1 extending from about residue 175 to about
 20 residue 230 (i.e., the EGF-like domain). For example, the fragment can extend from residue 177 to residue
 244 and may be prepared by recombinant techniques (rHRG β 1-177-244).

DNA encoding amino acid sequence variants of HRG is prepared by a variety of methods known in
 the art. These methods include, but are not limited to, isolation from a natural source (in the case of naturally
 occurring amino acid sequence variants) or preparation by oligonucleotide-mediated (or site-directed)
 25 mutagenesis, PCR mutagenesis, and cassette mutagenesis of an earlier prepared variant or a non-variant
 version of HRG. These techniques may utilize HRG nucleic acid (DNA or RNA), or nucleic acid
 complementary to HRG nucleic acid.

Oligonucleotide-mediated mutagenesis is a preferred method for preparing substitution, deletion,
 and insertion variants of HRG DNA. This technique is well known in the art as described by Adelman *et al.*
 30 *DNA*, 2: 183 (1983). Briefly, HRG DNA is altered by hybridizing an oligonucleotide encoding the desired
 mutation to a DNA template, where the template is the single-stranded form of a plasmid or bacteriophage
 containing the unaltered or native DNA sequence of HRG. After hybridization, a DNA polymerase is used
 to synthesize an entire second complementary strand of the template that will thus incorporate the
 oligonucleotide primer, and will code for the selected alteration in HRG DNA.

35 Generally, oligonucleotides of at least 25 nucleotides in length are used. An optimal
 oligonucleotide will have 12 to 15 nucleotides that are completely complementary to the template on either
 side of the nucleotide(s) coding for the mutation. This ensures that the oligonucleotide will hybridize
 properly to the single-stranded DNA template molecule. The oligonucleotides are readily synthesized using
 techniques known in the art such as that described by Crea *et al.* (*Proc. Natl. Acad. Sci. USA*, 75:
 40 5765, 1978).

Single-stranded DNA template may also be generated by denaturing double-stranded plasmid (or
 other) DNA using standard techniques.

For alteration of the native DNA sequence (to generate amino acid sequence variants, for example),
 the oligonucleotide is hybridized to the single-stranded template under suitable hybridization conditions. A
 45 DNA polymerizing enzyme, usually the Klenow fragment of DNA polymerase I, is then added to synthesize
 the complementary strand of the template using the oligonucleotide as a primer for synthesis. A

primers is designed to overlap the position of the mutation and to contain the mutation; the sequence of the other primer must be identical to a stretch of sequence of the opposite strand of the plasmid, but this sequence can be located anywhere along the plasmid DNA. It is preferred, however, that the sequence of the second primer is located within 200 nucleotides from that of the first, such that in the end the entire amplified region of DNA bounded by the primers can be easily sequenced. PCR amplification using a primer pair like the one just described results in a population of DNA fragments that differ at the position of the mutation specified by the primer, and possibly at other positions, as template copying is somewhat error-prone.

If the ratio of template to product material is extremely low, the vast majority of product DNA fragments incorporate the desired mutation(s). This product material is used to replace the corresponding region in the plasmid that served as PCR template using standard DNA technology. Mutations at separate positions can be introduced simultaneously by either using a mutant second primer, or performing a second PCR with different mutant primers and ligating the two resulting PCR fragments simultaneously to the vector fragment in a three (or more)-part ligation.

In a specific example of PCR mutagenesis, template plasmid DNA (1mg) is linearized by digestion with a restriction endonuclease that has a unique recognition site in the plasmid DNA outside of the region to be amplified. Of this material, 100 ng is added to a PCR mixture containing PCR buffer, which contains the four deoxynucleotide tri-phosphates and is included in the GENEAMP kits (obtained from Perkin-Elmer Cetus, Norwalk, CT and Emeryville, CA), and 25 pmole of each oligonucleotide primer, to a final volume of 50 µl. The reaction mixture is overlaid with 35 µl mineral oil. The reaction is denatured for 5 minutes at 100°C, placed briefly on ice, and then 1 µl *Thermus aquaticus* (Taq) DNA polymerase (5 units/ml, purchased from Perkin-Elmer Cetus, Norwalk, CT and Emeryville, CA) is added below the mineral oil layer. The reaction mixture is then inserted into a DNA Thermal Cycler (purchased from Perkin-Elmer Cetus) programmed as follows:

2 min. 55°C,
30 sec. 72°C, then 19 cycles of the following:
30 sec. 94°C,
30 sec. 55°C, and
30 sec. 72°C.

At the end of the program, the reaction vial is removed from the thermal cycler and the aqueous phase transferred to a new vial, extracted with phenol/chloroform (50:50:vol), and ethanol precipitated, and the DNA is recovered by standard procedures. This material is subsequently subjected to the appropriate treatments for insertion into a vector.

Another method for preparing variants, cassette mutagenesis, is based on the technique described by Wells *et al.* (*Gene*, 34: 315,1985). The starting material is the plasmid (or other vector) comprising HRG DNA to be mutated. The codon(s) in HRG DNA to be mutated are identified. There must be a unique restriction endonuclease site on each side of the identified mutation site(s). If no such restriction sites exist, they may be generated using the above-described oligonucleotide-mediated mutagenesis method to introduce them at appropriate locations in HRG DNA. After the restriction sites have been introduced into the plasmid, the plasmid is cut at these sites to linearize it. A double-stranded oligonucleotide encoding the sequence of the DNA between the restriction sites but containing the desired mutation(s) is synthesized using standard procedures. The two strands are synthesized separately and then hybridized together using standard techniques. This double-stranded oligonucleotide is referred to as the cassette. This cassette is designed to have 3' and 5' ends that are compatible with the ends of the linearized plasmid, such that it can be directly ligated to the plasmid. This plasmid now contains the mutated HRG DNA sequence.

C. Insertion of DNA into a Cloning Vehicle

The cDNA or genomic DNA encoding native or variant HRG is inserted into a replicable vector for further cloning (amplification of the DNA) or for expression. Many vectors are available, and selection of

encoding HRG is more complex than that of an exogenously replicated vector because restriction enzyme digestion is required to excise HRG DNA. DNA can be amplified by PCR and directly transfected into the host cells without any replication component.

(iii) Selection Gene Component

5 Expression and cloning vectors should contain a selection gene, also termed a selectable marker. This gene encodes a protein necessary for the survival or growth of transformed host cells grown in a selective culture medium. Host cells not transformed with the vector containing the selection gene will not survive in the culture medium. Typical selection genes encode proteins that (a) confer resistance to antibiotics or other toxins, e.g., ampicillin, neomycin, methotrexate, or tetracycline, (b) complement
10 auxotrophic deficiencies, or (c) supply critical nutrients not available from complex media, e.g., the gene encoding D-alanine racemase for *Bacilli*.

One example of a selection scheme utilizes a drug to arrest growth of a host cell. Those cells that are successfully transformed with a heterologous gene express a protein conferring drug resistance and thus survive the selection regimen. Examples of such dominant selection use the drugs neomycin (Southern *et al.*, *J. Molec. Appl. Genet.* 1: 327,1982), mycophenolic acid (Mulligan *et al.*, *Science* 209: 1422,1980) or
15 hygromycin (Sugden *et al.*, *Mol. Cell. Biol.* 5: 410-413,1985). The three examples given above employ bacterial genes under eukaryotic control to convey resistance to the appropriate drug G418 or neomycin (geneticin), xgpt (mycophenolic acid), or hygromycin, respectively.

Another example of suitable selectable markers for mammalian cells are those that enable the
20 identification of cells competent to take up HRG nucleic acid, such as dihydrofolate reductase (DHFR) or thymidine kinase. The mammalian cell transformants are placed under selection pressure which only the transformants are uniquely adapted to survive by virtue of having taken up the marker. Selection pressure is imposed by culturing the transformants under conditions in which the concentration of selection agent in the medium is successively changed, thereby leading to amplification of both the selection gene and the DNA
25 that encodes HRG. Amplification is the process by which genes in greater demand for the production of a protein critical for growth are reiterated in tandem within the chromosomes of successive generations of recombinant cells. Increased quantities of HRG are synthesized from the amplified DNA.

For example, cells transformed with the DHFR selection gene are first identified by culturing all of the transformants in a culture medium that contains methotrexate (Mtx), a competitive antagonist of DHFR.
30 An appropriate host cell when wild-type DHFR is employed is the Chinese hamster ovary (CHO) cell line deficient in DHFR activity, prepared and propagated as described by Urlaub and Chasin, *Proc. Natl. Acad. Sci. USA*, 77: 4216, 1980. The transformed cells are then exposed to increased levels of methotrexate. This leads to the synthesis of multiple copies of the DHFR gene, and, concomitantly, multiple copies of other DNA comprising the expression vectors, such as the DNA encoding HRG. This amplification technique can
35 be used with any otherwise suitable host, e.g., ATCC No. CCL61 CHO-K1, notwithstanding the presence of endogenous DHFR if, for example, a mutant DHFR gene that is highly resistant to Mtx is employed (EP 117,060). Alternatively, host cells (particularly wild-type hosts that contain endogenous DHFR) transformed or co-transformed with DNA sequences encoding HRG, wild-type DHFR protein, and another selectable marker such as aminoglycoside 3' phosphotransferase (APH) can be selected by cell growth in medium
40 containing a selection agent for the selectable marker such as an aminoglycosidic antibiotic, e.g., kanamycin, neomycin, or G418 (see U.S. Pat. No. 4,965,199).

A suitable selection gene for use in yeast is the *trp1* gene present in the yeast plasmid YRp7 (Stinchcomb *et al.*, *Nature*, 282: 39, 1979; Kingsman *et al.*, *Gene*, 7: 141, 1979; or Tschemper *et al.*, *Gene*, 10: 157, 1980). The *trp1* gene provides a selection marker for a mutant strain of yeast lacking the ability to
45 grow in tryptophan; for example, ATCC No. 44076 or PEP4-1 (Jones, *Genetics*, 85: 12, 1977). The presence of the *trp1* lesion in the yeast host cell genome then provides an effective environment for detecting

promoters. e.g., the actin promoter or an immunoglobulin promoter, from heat-shock promoters, and from the promoter normally associated with HRG sequence, provided such promoters are compatible with the host cell systems.

The early and late promoters of the SV40 virus are conveniently obtained as an SV40 restriction fragment that also contains the SV40 viral origin of replication (Fiers *et al.*, *Nature*, 273:113 (1978); Mulligan and Berg, *Science*, 209: 1422-1427 (1980); Pavlakis *et al.*, *Proc. Natl. Acad. Sci. USA*, 78: 7398-7402 (1981)). The immediate early promoter of the human cytomegalovirus is conveniently obtained as a *HindIII* E restriction fragment (Greenaway *et al.*, *Gene*, 18: 355-360 (1982)). A system for expressing DNA in mammalian hosts using the bovine papilloma virus as a vector is disclosed in U.S. 4,419,446. A modification of this system is described in U.S. 4,601,978. See also Gray *et al.*, *Nature*, 295: 503-508 (1982) on expressing cDNA encoding immune interferon in monkey cells; Reyes *et al.*, *Nature*, 297: 598-601 (1982) on expression of human b-interferon cDNA in mouse cells under the control of a thymidine kinase promoter from herpes simplex virus; Canaani and Berg, *Proc. Natl. Acad. Sci. USA*, 79: 5166-5170 (1982) on expression of the human interferon b1 gene in cultured mouse and rabbit cells; and Gorman *et al.*, *Proc. Natl. Acad. Sci. USA*, 79: 6777-6781 (1982) on expression of bacterial CAT sequences in CV-1 monkey kidney cells, chicken embryo fibroblasts, Chinese hamster ovary cells, HeLa cells, and mouse NIH-3T3 cells using the Rous sarcoma virus long terminal repeat as a promoter.

(v) Enhancer Element Component

Transcription of a DNA encoding HRG of this invention by higher eukaryotes is often increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting elements of DNA, usually about from 10-300 bp, that act on a promoter to increase its transcription. Enhancers are relatively orientation and position independent having been found 5' (Laimins *et al.*, *Proc. Natl. Acad. Sci. USA*, 78: 993, 1981) and 3' (Lusky *et al.*, *Mol. Cell Bio.*, 3: 1108, 1983) to the transcription unit, within an intron (Banerji *et al.*, *Cell*, 33: 729, 1983) as well as within the coding sequence itself (Osborne *et al.*, *Mol. Cell Bio.*, 4: 1293, 1984). Many enhancer sequences are now known from mammalian genes (globin, elastase, albumin, a-fetoprotein and insulin). Typically, however, one will use an enhancer from a eukaryotic cell virus. Examples include the SV40 enhancer on the late side of the replication origin (bp 100-270), the cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers (see also Yaniv, *Nature*, 297: 17-18 (1982)) on enhancing elements for activation of eukaryotic promoters. The enhancer may be spliced into the vector at a position 5' or 3' to HRG DNA, but is preferably located at a site 5' from the promoter.

(vi) Transcription Termination Component

Expression vectors used in eukaryotic host cells (yeast, fungi, insect, plant, animal, human, or nucleated cells from other multicellular organisms) will also contain sequences necessary for the termination of transcription and for stabilizing the mRNA. Such sequences are commonly available from the 5' and, occasionally 3' untranslated regions of eukaryotic or viral DNAs or cDNAs. These regions contain nucleotide segments transcribed as polyadenylated fragments in the untranslated portion of the mRNA encoding HRG. The 3' untranslated regions also include transcription termination sites.

Construction of suitable vectors containing one or more of the above listed components the desired coding and control sequences employs standard ligation techniques. Isolated plasmids or DNA fragments are cleaved, tailored, and religated in the form desired to generate the plasmids required.

For analysis to confirm correct sequences in plasmids constructed, the ligation mixtures are used to transform *E. coli* K12 strain 294 (ATCC 31,446) and successful transformants selected by ampicillin or tetracycline resistance where appropriate. Plasmids from the transformants are prepared, analyzed by restriction endonuclease digestion, and/or sequenced by the method of Messing *et al.*, *Nucleic Acids Res.* 9: 309 (1981) or by the method of Maxam *et al.*, *Methods in Enzymology* 65: 499 (1980).

Plant cell cultures of cotton, corn, potato, soybean, petunia, tomato, and tobacco can be utilized as hosts. Typically, plant cells are transfected by incubation with certain strains of the bacterium *Agrobacterium tumefaciens*, which has been previously manipulated to contain HRG DNA. During incubation of the plant cell culture with *A. tumefaciens*, the DNA encoding HRG is transferred to the plant cell host such that it is transfected, and will, under appropriate conditions, express HRG DNA. In addition, regulatory and signal sequences compatible with plant cells are available, such as the nopaline synthase promoter and polyadenylation signal sequences (Depicker *et al.*, *J. Mol. Appl. Gen.*, 1: 561 (1982)). In addition, DNA segments isolated from the upstream region of the T-DNA 780 gene are capable of activating or increasing transcription levels of plant-expressible genes in recombinant DNA-containing plant tissue (see EP 321,196, published 21 June 1989).

However, interest has been greatest in vertebrate cells, and propagation of vertebrate cells in culture (tissue culture) has become a routine procedure in recent years (*Tissue Culture*, Academic Press, Kruse and Patterson, editors (1973)). Examples of useful mammalian host cell lines are monkey kidney CV1 line transformed by SV40 (COS-7, ATCC CRL 1651); human embryonic kidney line (293 or 293 cells subcloned for growth in suspension culture, Graham *et al.*, *J. Gen. Virol.*, 36: 59, 1977); baby hamster kidney cells (BHK, ATCC CCL 10); Chinese hamster ovary cells/-DHFR (CHO, Urlaub and Chasin, *Proc. Natl. Acad. Sci. USA*, 77: 4216 (1980)); mouse sertoli cells (TM4, Mather, *Biol. Reprod.*, 23: 243-251 (1980)); monkey kidney cells (CV1 ATCC CCL 70); African green monkey kidney cells (VERO-76, ATCC CRL-1587); human cervical carcinoma cells (HELA, ATCC CCL 2); canine kidney cells (MDCK, ATCC CCL 34); buffalo rat liver cells (BRL 3A, ATCC CRL 1442); human lung cells (W138, ATCC CCL 75); human liver cells (Hep G2, HB 8065); mouse mammary tumor (MMT 060562, ATCC CCL51); TRI cells (Mather *et al.*, *Annals N.Y. Acad. Sci.*, 383: 44-68 (1982)); MRC 5 cells; FS4 cells; and a human hepatoma cell line (Hep G2). Preferred host cells are human embryonic kidney 293 and Chinese hamster ovary cells.

Host cells are transfected and preferably transformed with the above-described expression or cloning vectors of this invention and cultured in conventional nutrient media modified as appropriate for inducing promoters, selecting transformants, or amplifying the genes encoding the desired sequences.

Transfection refers to the taking up of an expression vector by a host cell whether or not any coding sequences are in fact expressed. Numerous methods of transfection are known to the ordinarily skilled artisan, for example, CaPO_4 and electroporation. Successful transfection is generally recognized when any indication of the operation of this vector occurs within the host cell.

Transformation means introducing DNA into an organism so that the DNA is replicable, either as an extrachromosomal element or by chromosomal integration. Depending on the host cell used, transformation is done using standard techniques appropriate to such cells. The calcium treatment employing calcium chloride, as described in section 1.82 of Sambrook *et al.*, *supra*, is generally used for prokaryotes or other cells that contain substantial cell-wall barriers. Infection with *Agrobacterium tumefaciens* is used for transformation of certain plant cells, as described by Shaw *et al.*, *Gene*, 23: 315 (1983) and WO 89/05859, published 29 June 1989. For mammalian cells without such cell walls, the calcium phosphate precipitation method described in sections 16.30-16.37 of Sambrook *et al.*, *supra*, is preferred. General aspects of mammalian cell host system transformations have been described by Axel in U.S. 4,399,216, issued 16 August 1983. Transformations into yeast are typically carried out according to the method of Van Solingen *et al.*, *J. Bact.*, 130: 946 (1977) and Hsiao *et al.*, *Proc. Natl. Acad. Sci. (USA)*, 76: 3829 (1979). However, other methods for introducing DNA into cells such as by nuclear injection, electroporation, or protoplast fusion may also be used.

E. Culturing the Host Cells

Prokaryotic cells used to produce HRG polypeptide of this invention are cultured in suitable media as described generally in Sambrook *et al.*, *supra*.

prepared against a native HRG polypeptide or against a synthetic peptide based on the DNA sequences provided herein as described further below.

G. Purification of The Heregulin Polypeptides

HRG is recovered from a cellular membrane fraction. Alternatively, a proteolytically cleaved or a truncated expressed soluble HRG fragment or subdomain are recovered from the culture medium as a soluble polypeptide. A HRG is recovered from host cell lysates when directly expressed without a secretory signal.

When HRG is expressed in a recombinant cell other than one of human origin, HRG is completely free of proteins or polypeptides of human origin. However, it is desirable to purify HRG from recombinant cell proteins or polypeptides to obtain preparations that are substantially homogeneous as to HRG. As a first step, the culture medium or lysate is centrifuged to remove particulate cell debris. The membrane and soluble protein fractions are then separated. HRG is then be purified from both the soluble protein fraction (requiring the presence of a protease) and from the membrane fraction of the culture lysate, depending on whether HRG is membrane bound. The following procedures are exemplary of suitable purification procedures: fractionation on immunoaffinity or ion-exchange columns; ethanol precipitation; reverse phase HPLC; chromatography on silica, heparin SEPHAROSE or on a cation exchange resin such as DEAE; chromatofocusing; SDS-PAGE; ammonium sulfate precipitation; and gel filtration using, for example, SEPHADEX G-75.

HRG variants in which residues have been deleted, inserted or substituted are recovered in the same fashion as the native HRG, taking account of any substantial changes in properties occasioned by the variation. For example, preparation of a HRG fusion with another protein or polypeptide, e.g., a bacterial or viral antigen, facilitates purification; an immunoaffinity column containing antibody to the antigen can be used to adsorb the fusion. Immunoaffinity columns such as a rabbit polyclonal anti-HRG column can be employed to absorb HRG variant by binding it to at least one remaining immune epitope. A protease inhibitor such as phenylmethylsulfonylfluoride (PMSF) also may be useful to inhibit proteolytic degradation during purification, and antibiotics may be included to prevent the growth of adventitious contaminants. One skilled in the art will appreciate that purification methods suitable for native HRG may require modification to account for changes in the character of HRG variants or upon expression in recombinant cell culture.

H. Covalent Modifications of HRG

Covalent modifications of HRG polypeptides are included within the scope of this invention. Both native HRG and amino acid sequence variants of HRG optionally are covalently modified. One type of covalent modification included within the scope of this invention is a HRG polypeptide fragment. HRG fragments, such as HRG-GDF, having up to about 40 amino acid residues are conveniently prepared by chemical synthesis, or by enzymatic or chemical cleavage of the full-length HRG polypeptide or HRG variant polypeptide. Other types of covalent modifications of HRG or fragments thereof are introduced into the molecule by reacting targeted amino acid residues of HRG or fragments thereof with an organic derivatizing agent that is capable of reacting with selected side chains or the N- or C-terminal residues.

Cysteiny residues most commonly are reacted with α -haloacetates (and corresponding amines), such as chloroacetic acid or chloroacetamide, to give carboxymethyl or carboxyamidomethyl derivatives. Cysteiny residues also are derivatized by reaction with bromotrifluoroacetone, α -bromo- β -(5-imidozoyl)propionic acid, chloroacetyl phosphate, N-alkylmaleimides, 3-nitro-2-pyridyl disulfide, methyl 2-pyridyl disulfide, p-chloromercuribenzoate, 2-chloromercuri-4-nitrophenol, or chloro-7-nitrobenzo-2-oxa-1,3-diazole.

Histidyl residues are derivatized by reaction with diethylpyrocarbonate at pH 5.5-7.0 because this agent is relatively specific for the histidyl side chain. Para-bromophenacyl bromide also is useful; the reaction is preferably performed in 0.1M sodium cacodylate at pH 6.0.

asparagine-X-serine and asparagine-X-threonine, where X is any amino acid except proline, are the recognition sequences for enzymatic attachment of the carbohydrate moiety to the asparagine side chain. Thus, the presence of either of these tri-peptide sequences in a polypeptide creates a potential glycosylation site. O-linked glycosylation refers to the attachment of one of the sugars N-acetylgalactosamine, galactose, or xylose, to a hydroxyamino acid, most commonly serine or threonine, although 5-hydroxyproline or 5-hydroxylysine may also be used.

Glycosylation sites are added to HRG by altering its amino acid sequence to contain one or more of the above-described tri-peptide sequences (for N-linked glycosylation sites). The alteration may also be made by the addition of, or substitution by, one or more serine or threonine residues to HRG (for O-linked glycosylation sites). For ease, HRG is preferably altered through changes at the DNA level, particularly by mutating the DNA encoding HRG at preselected bases such that codons are generated that will translate into the desired amino acids.

Chemical or enzymatic coupling of glycosides to HRG increases the number of carbohydrate substituents. These procedures are advantageous in that they do not require production of the polypeptide in a host cell that is capable of N- and O-linked glycosylation. Depending on the coupling mode used, the sugar(s) may be attached to (a) arginine and histidine, (b) free carboxyl groups, (c) free sulfhydryl groups such as those of cysteine, (d) free hydroxyl groups such as those of serine, threonine, or hydroxyproline, (e) aromatic residues such as those of phenylalanine, tyrosine, or tryptophan, or (f) the amide group of glutamine. These methods are described in WO 87/05330, published 11 September 1987, and in Aplin and Wriston (*CRC Crit. Rev. Biochem.*, pp. 259-306 (1981)).

Carbohydrate moieties present on an HRG also are removed chemically or enzymatically. Chemical deglycosylation requires exposure of the polypeptide to the compound trifluoromethanesulfonic acid, or an equivalent compound. This treatment results in the cleavage of most or all sugars except the linking sugar (N-acetylglucosamine or N-acetylgalactosamine), while leaving the polypeptide intact. Chemical deglycosylation is described by Hakimuddin *et al.* (*Arch. Biochem. Biophys.*, 259:52 (1987)) and by Edge *et al.* (*Anal. Biochem.*, 118:131 (1981)). Carbohydrate moieties are removed from HRG by a variety of endo- and exo- glycosidases as described by Thotakura *et al.* (*Meth. Enzymol.*, 138:350 (1987)).

Glycosylation also is suppressed by tunicamycin as described by Duskin *et al.* (*J. Biol. Chem.*, 257:3105 (1982)). Tunicamycin blocks the formation of protein-N-glycoside linkages.

HRG may also be modified by linking HRG to various nonproteinaceous polymers, e.g., polyethylene glycol, polypropylene glycol or polyoxyalkylenes, in the manner set forth in U.S. 4,640,835; 4,496,689; 4,301,144; 4,670,417; 4,791,192 or 4,179,337.

One preferred way to increase the *in vivo* circulating half life of non-membrane bound HRG is to conjugate it to a polymer that confers extended half-life, such as polyethylene glycol (PEG). (Maxfield, *et al.*, *Polymer* 16,505-509 (1975); Bailey, F. E., *et al.*, in *Nonionic Surfactants* (Schick, M. J., ed.) pp.794-821, 1967); (Abuchowski, A. *et al.*, *J. Biol. Chem.* 252, 3582-3586, 1977; Abuchowski, A. *et al.*, *Cancer Biochem. Biophys.* 7, 175-186, 1984); (Katre, N.V. *et al.*, *Proc. Natl. Acad. Sci.*, 84, 1487-1491, 1987; Goodson, R. *et al.* *Bio Technology*, 8, 343-346, 1990). Conjugation to PEG also has been reported to have reduced immunogenicity and toxicity (Abuchowski, A. *et al.*, *J. Biol. Chem.*, 252, 3578-3581, 1977).

HRG may also be entrapped in microcapsules prepared, for example, by coacervation techniques or by interfacial polymerization (for example, hydroxymethylcellulose or gelatin-microcapsules and poly-(methylmethacrylate) microcapsules, respectively), in colloidal drug delivery systems (for example, liposomes, albumin microspheres, microemulsions, nano-particles and nanocapsules), or in macroemulsions. Such techniques are disclosed in *Remington's Pharmaceutical Sciences*, 16th edition, Osol, A., Ed., (1980).

Those skilled in the art will be capable of screening variants in order to select the optimal variant for the purpose intended. For example, a change in the immunological character of HRG, such as a change in affinity for a given antigen or for the HER2 receptor, is measured by a competitive-type immunoassay using

Culture medium in which hybridoma cells are growing is assayed for production of monoclonal antibodies directed against the antigen. Preferably, the binding specificity of monoclonal antibodies produced by hybridoma cells is determined by immunoprecipitation or by an *in vitro* binding assay, such as radioimmunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA).

5 The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson *et al.*, *Anal. Biochem.*, 107:220 (1980).

After hybridoma cells are identified that produce antibodies of the desired specificity, affinity, and/or activity, the clones may be subcloned by limiting dilution procedures and grown by standard methods (Goding, *Monoclonal Antibodies: Principles and Practice*, pp.59-103 (Academic Press, 1986)). Suitable
10 culture media for this purpose include, for example, D-MEM or RPMI-1640 medium. In addition, the hybridoma cells may be grown *in vivo* as ascites tumors in an animal.

The monoclonal antibodies secreted by the subclones are suitably separated from the culture medium, ascites fluid, or serum by conventional immunoglobulin purification procedures such as, for example, protein A-SEPHAROSE, hydroxylapatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

15 DNA encoding the monoclonal antibodies is readily isolated and sequenced using conventional procedures (e.g., by using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of the monoclonal antibodies). The hybridoma cells serve as a preferred source of such DNA. Once isolated, the DNA may be placed into expression vectors, which are then transfected into host cells such as *E. coli* cells, simian COS cells, Chinese hamster ovary (CHO) cells, or myeloma cells that
20 do not otherwise produce immunoglobulin protein, to obtain the synthesis of monoclonal antibodies in the recombinant host cells. Recombinant production of antibodies will be described in more detail below.

Hybridoma cell lines producing antibodies are identified by screening the culture supernatants for antibody which binds to HER2/HER3 receptors. This is routinely accomplished by conventional immunoassays using soluble receptor preparations or by FACS using cell-bound receptor and labeled
25 candidate antibody. Agonist antibodies are preferably antibodies which stimulate autophosphorylation in the HRG tyrosine autophosphorylation assay described above.

The hybrid cell lines can be maintained in culture *in vitro* in cell culture media. The cell lines of this invention can be selected and/or maintained in a composition comprising the continuous cell line in hypoxanthine-aminopterin thymidine (HAT) medium. In fact, once the hybridoma cell line is established, it
30 can be maintained on a variety of nutritionally adequate media. Moreover, the hybrid cell lines can be stored and preserved in any number of conventional ways, including freezing and storage under liquid nitrogen. Frozen cell lines can be revived and cultured indefinitely with resumed synthesis and secretion of monoclonal antibody. The secreted antibody is recovered from tissue culture supernatant by conventional methods such as precipitation, ion exchange chromatography, affinity chromatography, or the like. The
35 antibodies described herein are also recovered from hybridoma cell cultures by conventional methods for purification of IgG or IgM as the case may be that heretofore have been used to purify these immunoglobulins from pooled plasma, e.g., ethanol or polyethylene glycol precipitation procedures.

Human antibodies may be used and are preferable. Such antibodies can be obtained by using human hybridomas (Cote *et al.*, *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, p. 77 (1985)).
40 Chimeric antibodies, Cabilly *et al.*, U.S. 4,816,567, (Morrison *et al.*, *Proc. Natl. Acad. Sci.*, 81:6851 (1984); Neuberger *et al.*, *Nature* 312:604 (1984); Takeda *et al.*, *Nature* 314:452 (1985)) containing a murine anti-HER2/HER3 variable region and a human constant region of appropriate biological activity (such as ability to activate human complement and mediate ADCC) are within the scope of this invention, as are humanized antibodies produced by conventional CDR-grafting methods (Riechmann *et al.*, *Nature* 332:333-327(1988);
45 EP 0328404 A1; EP 02394000 A2).

Techniques for creating recombinant DNA versions of the antigen-binding regions of antibody molecules (Fab or variable regions fragments) which bypass the generation of monoclonal antibodies are also

Therapeutic formulations of HRG or agonist antibody are prepared for storage by mixing the HRG protein having the desired degree of purity with optional physiologically acceptable carriers, excipients, or stabilizers (*Remington's Pharmaceutical Sciences, supra*), in the form of lyophilized cake or aqueous solutions. Acceptable carriers, excipients or stabilizers are nontoxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, and other organic acids; antioxidants including ascorbic acid; low molecular weight (less than about 10 residues) polypeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids such as glycine, glutamine, asparagine, arginine or lysine; monosaccharides, disaccharides, and other carbohydrates including glucose, mannose, or dextrans; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; salt-forming counterions such as sodium; and/or nonionic surfactants such as TWEEN, PLURONICS or polyethylene glycol (PEG).

HRG or agonist antibody to be used for *in vivo* administration must be sterile. This is readily accomplished by filtration through sterile filtration membranes, prior to or following lyophilization and reconstitution. The HRG or antibody ordinarily will be stored in lyophilized form or in solution.

Therapeutic HRG or antibody compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

The route of HRG or antibody administration is in accord with known methods, e.g., injection or infusion by intravenous, intraperitoneal, intracerebral, intramuscular, intraocular, intraarterial, powder or liquid aerosol administration to the nose or lung or intralesional routes, or by sustained release systems as noted below. The HRG ligand may be administered continuously by infusion or by bolus injection. An agonist antibody is preferably administered in the same fashion, or by administration into the blood stream or lymph.

The HRG, HRG variant or fragment and agonist antibodies may be spray dried or spray freeze dried using known techniques (Yeo et al, *Biotech. and Bioeng.*, 41:341-346 (1993); Gombotz et al, PCT/US90/02421).

Suitable examples of sustained-release preparations include semipermeable matrices of solid hydrophobic polymers containing the protein, which matrices are in the form of shaped articles, e.g. films, or microcapsules. Examples of sustained-release matrices include polyesters, hydrogels (e.g., poly(2-hydroxyethyl-methacrylate) as described by Langer et al., *J. Biomed. Mater. Res.*, 15: 167-277 (1981) and Langer, *Chem. Tech.*, 12: 98-105 (1982) or poly(vinylalcohol)), polylactides (U.S. 3,773,919, EP 58,481), copolymers of L-glutamic acid and gamma ethyl-L-glutamate (Sidman et al., *Biopolymers*, 22: 547-556 (1983)), non-degradable ethylene-vinyl acetate (Langer et al., *supra*), degradable lactic acid-glycolic acid copolymers such as the LUPRON DEPOT (injectable microspheres composed of lactic acid-glycolic acid copolymer and leuprolide acetate), and poly-D-(-)-3-hydroxybutyric acid (EP 133,988). While polymers such as ethylene-vinyl acetate and lactic acid-glycolic acid enable release of molecules for over 100 days, certain hydrogels release proteins for shorter time periods. When encapsulated proteins remain in the body for a long time, they may denature or aggregate as a result of exposure to moisture at 37°C, resulting in a loss of biological activity and possible changes in immunogenicity. Rational strategies can be devised for protein stabilization depending on the mechanism involved. For example, if the aggregation mechanism is discovered to be intermolecular S-S bond formation through thio-disulfide interchange, stabilization may be achieved by modifying sulfhydryl residues, lyophilizing from acidic solutions, controlling moisture content, using appropriate additives, and developing specific polymer matrix compositions.

Sustained-release HRG or antibody compositions also include liposomally entrapped HRG or antibody. Liposomes containing HRG or antibody are prepared by methods known *per se*: DE 3,218,121; Epstein et al., *Proc. Natl. Acad. Sci. USA*, 82: 3688-3692 (1985); Hwang et al., *Proc. Natl. Acad. Sci. USA*,

L⁶³⁶ HER3-(TR)-DKTH²²⁴_{VH}; for HER4, G⁶⁴⁰ HER4-(TR)-DKTH²²⁴_{VH}. The conserved TR sequence is derived from the MI I site. The final expression constructs were in a pRK-type plasmid backbone wherein eukaryotic expression is driven by a CMV promoter (Gorman *et al.*, *DNA Prot. Eng. Tech.* 2:3-10 (1990)).

To obtain protein for *in vitro* experiments, adherent HEK-293 cells were transfected with the appropriate expression plasmids using standard calcium phosphate methods (Gorman *et al.*, *supra* and Huang *et al.*, *Nucleic Acids Res.* 18:937-947 (1990)). Serum-containing media was replaced with serum-free media 15 hours post-transfection and the transfected cells incubated for 5-7 days. The resulting conditioned media was harvested and passed through Protein A columns (1 mL Pharmacia HiTrap™). Purified IgG fusions were eluted with 0.1 M citric acid (pH 4.2) into tubes containing 1 M Tris pH 9.0. The eluted proteins were subsequently dialyzed against PBS and concentrated using Centri-prep-30 filters (Amicon). Glycerol was added to a final concentration of 25% and the material stored at -20 C. Concentrations of material were determined via a Fc-ELISA.

Cell Culture: Human breast cancer cell lines MDA-MB-175, MDA-MB-231, SK-BR-3 and MCF7 were obtained from the American Type Culture Collection and maintained in a 50:50 mixture of F12 Ham's and Dulbecco's modified Eagle medium (DMEM), supplemented with 10% heat inactivated FBS, 2 mM glutamine and 10% penicillin-streptomycin.

Generation and Characterization of cDNA Library: Total RNA was purified from MDA-MB-175 cells using the guanidinium isothiocyanate-cesium chloride procedure (Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual* (New York: Cold Spring Harbor Laboratory Press, (1989)). Poly (A)⁺ RNA was isolated using oligo (dT) Dynabeads (DYNAL) as recommended by the supplier. First and second strand syntheses were performed using a Gibco BRL cDNA synthesis kit. λ gt10 cDNA recombinants were generated when a cDNA cloning system from Amersham was used. *In vitro* packaging was performed using Gigapack II packaging extract (Stratagene). PstI-XhoI HRG β 3 cDNA fragment (nt 144-618) was labeled by random priming and 1 x 10⁶ plaques were screened. Positive clones were confirmed and purified by secondary and tertiary screening. Phage DNA was isolated as a BamHI fragment and subcloned into the corresponding site of pBluescript SK⁻. Clone 5 was completely sequenced using the Sequenase version 2.0 DNA sequencing kit (United States Biochemicals, Inc.). Both strands were sequenced.

Bacterial Expression System: A cDNA fragment of clone 5 (nt 1690- 2722) was subcloned into the pET-32 TRX fusion vector (Novagen). This BglII-BglII fragment was inserted into the BamHI site of the pET32a plasmid. The trxy-HRG (amino acids 455-768) protein expression in *E.coli* was induced as recommended by the supplier.

Purification of Recombinant γ -HRG: *E.coli* cells expressing trxy-HRG were collected and suspended at 9 ml/g in 50 mM Tris HCL pH 8. Lysozyme was added to a final concentration of 0.2 mg/ml and the solution was stirred on ice for 1 hr. Dnase I (10 μ g/ml) and MgCl₂ (4 mM) were added. The solution was then sonicated for 30 min and cell pellets collected afterwards. The pellet fraction was dissolved at 250 ml/g in 6 M Gdn HCL, 0.1 M Tris HCL, pH 8.8. Solubilized proteins were sulfitolyzed by adding 1/10 volume of 1 M Na₂SO₃ and 1/10 volume of 0.2 M Na₂S₄O₆. The reaction was allowed to proceed for 1.5 hours at room temperature and protein was purified by gel filtration chromatography using a High Load Superdex™ 75 prep grade column (Pharmacia). Refolding was initiated by the addition of 1 mM cysteine, and 10 mM methionine was added as an antioxidant and incubated overnight at room temperature. Protein concentration was determined by quantitative amino acid analysis.

Northern and Southern Hybridization: Total RNA was isolated by the method of Chomczynski *et al.* *Anal Biochem.* 162:156-159 (1987). Poly (A)⁺ was isolated using oligo d(T) cellulose columns (Qiagen) as recommended by the supplier. RNA was denatured and size fractionated in a 0.8% formaldehyde/ 1%

seeded in T175 flasks and grown until reaching 70-80% confluency ($\sim 2.5 \times 10^7$ cells/ flask). Subsequently, cells were washed with PBS and grown in serum free F12/ DMEM medium for 3-4 days. Medium was then collected, filtered and concentrated using an ultrafiltration cell with YM10 Diaflo ultrafiltration membranes (Amicon). γ -HRG was visualized in conditioned medium of MDA-MB-175 cells by Western blot analysis under non reducing conditions. γ -HRG was partially purified by HPLC using a C4 reverse phase column. CHO expressed full length HRG β 1 (lane 1) and semi pure γ -HRG (lane 2) were electrophoresed, blot was probed with HER2/HER4 IgG heterodimers and Western blot was developed. A ~ 64 kDa band could be seen in the lane containing partial purified supernatant whereas CHO expressed full length HRG β 1 migrated as a 45 kDa protein.

Cell Proliferation Assay with Crystal Violet: Tumor cell lines were plated in 96 well plates at following densities: 2×10^4 cells/well for MDA-MB-175 and 1×10^4 cells/well for SK-BR-3. The media contained 1% FBS and cells were allowed to adhere for 2 hours. Monoclonal antibodies, immunoadhesions (10 μ g/ml) or media alone were added and the cells were incubated for 2 hours at 37°C. rHRG β 1₁₇₇₋₂₄₄ was added at a final concentration of 1 nM, or 100 nM for neutralising the immunoadhesion, and the cells were incubated for 4 days. Monolayers were washed with PBS and stained/fixed with 0.5% crystal violet. Plates were air dried, the dye was eluted with 0.1 M sodium citrate (pH 4.2) in ethanol (50:50) and the absorbance was measured at 540 nm.

Isolation and sequence analysis of γ -HRG: To characterize the heregulin transcript in MDA-MB-175 cells, a λ gt 10 cDNA library was constructed with mRNA derived from this cell line. The library was screened with a cDNA probe corresponding to the EGF-like domain and part of the N-terminal sequence of HRG β 3. Various clones were identified. One of the clones which appeared to contain the full length cDNA sequence was isolated and sequenced. Fig.7A-7D shows the nucleotide sequence and the predicted amino acid sequence of γ -HRG. The single open reading frame of 2303 bp starts with an ATG codon at nt 334. This start codon lies in a nucleotide sequence context, which is known to be a potential translation initiation site (Kozak, *Nucleic Acid Research* 15:8125-8148 (1987)). Several termination codons were found upstream of the initiation codon. The stop codon TAG at nt 2637 is followed by the 3' noncoding sequence, which is identical to other HRG isoform sequences and includes a polyadenylation signal followed by an A-rich region. The open reading frame encodes a protein of 768 amino acid residues with a calculated molecular mass of 84.2 kDa.

(d) The selection of HRG- β 1 variants containing residues corresponding to the minimal EGF-like domain (HRG- β 1 177-228) was conducted using monovalent phage display. For these variants, residue numbers also are expressed, in parentheses, in terms of the position of the residue in the minimal EGF-like domain (i.e., HRG- β 1 EGF 1-52).

Variants of HRG- β 1 EGF were prepared and selected for binding to HER-3-Ig using monovalent phage display, according to the method of Bass *et al.*, *Proteins* 8:309-314 (1990). As discussed in detail below, an HRG- β 1 EGF phagemid vector was prepared, in which HRG- β 1 EGF was fused to a C-terminal fragment of the M13 coat protein pIII. Kunkel mutagenesis was performed to introduce stop codons into this vector at sites selected for randomization. This step ensures that the starting vector is incapable of expressing the wild-type polypeptide. Stretches of four to six residues per library were randomized in a linear fashion, except for the six cysteines, Phe189 (HRG- β 1 EGF Phe13) and the two most C-terminal residues. Phe189 was not altered because this residue is conserved as an aromatic residue in EGF and TGF- α and forms a stacking interaction with Tyr208 (HRG- β 1 EGF Tyr32) Jacobsen *et al.*, *Biochemistry* 35:3402-

Table 4: Library C Variants

		Position in HRG- β 1				
Variant No.		191	192	193	194	195
Wild-type		V	N	G	G	E
1, 2, 4, 5, 7-12	
3		V
6		Q

Table 5: Library D Variants

		Position in HRG- β 1				
Variant No.		197	198	199	200	201
Wild-type		F	M	V	K	D
1*, 2*, 8*, 12*		Y	K	.	R	I
3		.	R	.	.	T
4, 5, 7, 9		Y	R	.	.	T
6		Y	.	I	.	Y
10		Y	.	.	.	T
11		M	R	.	R	T

*Variant also contained Met226Ile.

Table 6: Library E Variants

		Position in HRG- β 1				
Variant No.		205	206	207	208	209
Wild-type		P	S	R	Y	L
1		T	P	Y	L	M
2, 4		Y	G	Y	L	M
3*		Y	R	Y	R	M
5, 12		T	H	Y	R	G
6		T	H	Y	R	M
7*		Y	K	Y	R	M
8, 9		T	K	Y	R	G
10		Y	K	Y	R	.

*Variant also contained Met226Ile.

Table 7: Library G Variants

		Position in HRG- β 1					
Variant No.		211	212	213	214	215	216
Wild-type		K	C	P	N	E	F
1, 5, 6, 10, 12		R	.	S	L	.	.
2		R	.	S	E	.	.
3		.	.	.	K	.	M
4		R	.	T	V	.	Y
7, 8		R	.	T	V	.	Y
9		.	.	N	S	.	.
11		R	.	K	K	.	.

Example 2 - HER2/HER3 expression in embryonic rat lung

Rat lungs were microdissected from rat embryos on embryologic day (E) 16, 18, 20, and post-natal (P) days 7, 14, and adult. Isolated lung tissue was homogenized in a standard protease inhibitor buffer, and equal protein amounts subjected to SDS-PAGE (4-20%), blotted to nitrocellulose and identified with specific antibodies (HER2, HER3, HER4-Santa Cruz Biological, San Jose, CA. and HRG, 3G11, Genentech, Inc.) using chemiluminescent techniques.

Analysis of the blot indicates that HER2 is expressed at high levels throughout development in utero, appears to peak on E18 and declines after birth to lower adult levels. HER3 expression peaks at E18,

containing overhangs. The fragment was inserted into phagemid display vector pam-g3 by restriction digestion at the same sites to generate construct pHRG2-g3 (177-244). pam-g3 was a derivative of phGHam-g3, which was designed for phage display of human growth hormone (hGH) and was described in Lowman et al., *Biochemistry* 30:10832-10838 (1991). pam-g3 was produced by removing the hGH gene present in phGHam-g3 and replacing this gene with a stuffer fragment, which provides space for cleavage at the restriction sites used for cloning. The HRG- β 1 fragment was attached to residue 247 of pIII.

The HRG- β 1 EGF-like domain expressed from the above-described construct is designated by removing the "p" and the "-g3" that appear in the name of the construct. Thus, the HRG- β 1 EGF-like domain expressed from the pHRG2-g3 construct is designated "HRG2."

The domain was displayed monovalently on phage as a pIII fusion protein, as described by Bass et al., *Proteins* 8:309-314 (1990).

Similarly, variants HRG- β 1₁₄₇₋₂₂₇, HRG- β 1₁₄₇₋₂₄₄, and HRG- β 1₁₇₇₋₂₂₇ were prepared and expressed as described above.

Example 6 - rHRG β 1₁₇₇₋₂₄₄ causes accelerated lung development

Exogenous rHRG β 1₁₇₇₋₂₄₄ caused accelerated lung development *in vitro*. To determine if the expressed HER2/HER3 receptors were functional and the role of HRG stimulation during lung development *in vitro*, rHRG β 1₁₇₇₋₂₄₄ was added to the *in vitro* culture at 10 nM. Tissue was harvested at D5, snap frozen, and 5 micron sections were cut for analysis.

The morphology, in comparison to the untreated control specimens was grossly different. There was marked proliferation of the epithelium. Air spaces which are typically lined with a single cell layer now had a 2-3 cell thickness. The changes were dose dependent with more epithelial cell response with higher concentrations. HER2 and HER3 were still identifiable in the epithelium only.

Example 7 - Human lung differentiation

Differentiation of human lung epithelial cells occurs after HRG treatment. Differentiation was measured by Surfactant Protein A (SPA) production. All sections were stained for SPA. Human lung explant stained for SPA with stain localizing in epithelial cells of the prealveolar ducts. Human lung explant exposed to 10 nM HRG showed an effect on SPA production. As a differentiation control, a lung explant was exposed to 1mM dibutyryl cAMP. A negative control was also run.

17. The method of Claim 16, wherein the activating ligand is a heregulin (HRG) polypeptide, HRG variant, HRG agonist antibody or fragment thereof capable of binding to the HER2, HER3 and/or HER4 receptor.

18. A method of treating respiratory distress or emphysema, comprising administering to a patient in need thereof an effective amount of an isolated HER2, HER3 and/or HER4 activating ligand.

19. The method of Claim 18, wherein the activating ligand is a heregulin (HRG) polypeptide, HRG variant, HRG agonist antibody or fragment thereof capable of binding to the HER2, HER3 and/or HER4 receptor.

20. A method, comprising the steps of:

- (a) obtaining a normal epithelial cell sample from a mammal;
- (b) contacting the sample with a ligand which activates HER2, HER3, HER4 or a combination thereof to induce growth and/or proliferation of epithelial cells in the sample and to obtain an expanded sample; and
- (c) re-introducing the expanded sample into the mammal.

1 / 29

GG	GCG	CGA	GCG	CCT	CAG	CGC	GGC	CGC	TCG	CTC	TCC	CCC	38
	Ala	Arg	Ala	Pro	Gln	Arg	Gly	Arg	Ser	Leu	Ser	Pro	
	1				5					10			
TCG	AGG	GAC	AAA	CTT	TTC	CCA	AAC	CCG	ATC	CGA	GCC	CTT	77
Ser	Arg	Asp	Lys	Leu	Phe	Pro	Asn	Pro	Ile	Arg	Ala	Leu	
		15					20					25	
GGA	CCA	AAC	TCG	CCT	GCG	CCG	AGA	GCC	GTC	CGC	GTA	GAG	116
Gly	Pro	Asn	Ser	Pro	Ala	Pro	Arg	Ala	Val	Arg	Val	Glu	
				30					35				
CGC	TCC	GTC	TCC	GGC	GAG	ATG	TCC	GAG	CGC	AAA	GAA	GGC	155
Arg	Ser	Val	Ser	Gly	Glu	Met	Ser	Glu	Arg	Lys	Glu	Gly	
	40					45					50		
AGA	GGC	AAA	GGG	AAG	GGC	AAG	AAG	AAG	GAG	CGA	GGC	TCC	194
Arg	Gly	Lys	Gly	Lys	Gly	Lys	Lys	Lys	Glu	Arg	Gly	Ser	
			55					60					
GGC	AAG	AAG	CCG	GAG	TCC	GCG	GCG	GGC	AGC	CAG	AGC	CCA	233
Gly	Lys	Lys	Pro	Glu	Ser	Ala	Ala	Gly	Ser	Gln	Ser	Pro	
	65				70					75			
GCC	TTG	CCT	CCC	CGA	TTG	AAA	GAG	ATG	AAA	AGC	CAG	GAA	272
Ala	Leu	Pro	Pro	Arg	Leu	Lys	Glu	Met	Lys	Ser	Gln	Glu	
		80					85					90	
TCG	GCT	GCA	GGT	TCC	AAA	CTA	GTC	CTT	CGG	TGT	GAA	ACC	311
Ser	Ala	Ala	Gly	Ser	Lys	Leu	Val	Leu	Arg	Cys	Glu	Thr	
				95					100				
AGT	TCT	GAA	TAC	TCC	TCT	CTC	AGA	TTC	AAG	TGG	TTC	AAG	350
Ser	Ser	Glu	Tyr	Ser	Ser	Leu	Arg	Phe	Lys	Trp	Phe	Lys	
	105					110					115		
AAT	GGG	AAT	GAA	TTG	AAT	CGA	AAA	AAC	AAA	CCA	CAA	AAT	389
Asn	Gly	Asn	Glu	Leu	Asn	Arg	Lys	Asn	Lys	Pro	Gln	Asn	
			120					125					
ATC	AAG	ATA	CAA	AAA	AAG	CCA	GGG	AAG	TCA	GAA	CTT	CGC	428
Ile	Lys	Ile	Gln	Lys	Lys	Pro	Gly	Lys	Ser	Glu	Leu	Arg	
	130				135					140			
ATT	AAC	AAA	GCA	TCA	CTG	GCT	GAT	TCT	GGA	GAG	TAT	ATG	467
Ile	Asn	Lys	Ala	Ser	Leu	Ala	Asp	Ser	Gly	Glu	Tyr	Met	
		145				150						155	
TGC	AAA	GTG	ATC	AGC	AAA	TTA	GGA	AAT	GAC	AGT	GCC	TCT	506
Cys	Lys	Val	Ile	Ser	Lys	Leu	Gly	Asn	Asp	Ser	Ala	Ser	
				160					165				

FIG. 1A

2 / 29

GCC	AAT	ATC	ACC	ATC	GTG	GAA	TCA	AAC	GAG	ATC	ATC	ACT	545
Ala	Asn	Ile	Thr	Ile	Val	Glu	Ser	Asn	Glu	Ile	Ile	Thr	
170						175						180	
GGT	ATG	CCA	GCC	TCA	ACT	GAA	GGA	GCA	TAT	GTG	TCT	TCA	584
Gly	Met	Pro	Ala	Ser	Thr	Glu	Gly	Ala	Tyr	Val	Ser	Ser	
			185					190					
GAG	TCT	CCC	ATT	AGA	ATA	TCA	GTA	TCC	ACA	GAA	GGA	GCA	623
Glu	Ser	Pro	Ile	Arg	Ile	Ser	Val	Ser	Thr	Glu	Gly	Ala	
195					200					205			
AAT	ACT	TCT	TCA	TCT	ACA	TCT	ACA	TCC	ACC	ACT	GGG	ACA	662
Asn	Thr	Ser	Ser	Ser	Thr	Ser	Thr	Ser	Thr	Thr	Gly	Thr	
		210					215					220	
AGC	CAT	CTT	GTA	AAA	TGT	GCG	GAG	AAG	GAG	AAA	ACT	TTC	701
Ser	His	Leu	Val	Lys	Cys	Ala	Glu	Lys	Glu	Lys	Thr	Phe	
				225					230				
TGT	GTG	AAT	GGA	GGG	GAG	TGC	TTC	ATG	GTG	AAA	GAC	CTT	740
Cys	Val	Asn	Gly	Gly	Glu	Cys	Phe	Met	Val	Lys	Asp	Leu	
235						240					245		
TCA	AAC	CCC	TCG	AGA	TAC	TTG	TGC	AAG	TGC	CAA	CCT	GGA	779
Ser	Asn	Pro	Ser	Arg	Tyr	Leu	Cys	Lys	Cys	Gln	Pro	Gly	
			250					255					
TTC	ACT	GGA	GCA	AGA	TGT	ACT	GAG	AAT	GTG	CCC	ATG	AAA	818
Phe	Thr	Gly	Ala	Arg	Cys	Thr	Glu	Asn	Val	Pro	Met	Lys	
260					265					270			
GTC	CAA	AAC	CAA	GAA	AAG	GCG	GAG	GAG	CTG	TAC	CAG	AAG	857
Val	Gln	Asn	Gln	Glu	Lys	Ala	Glu	Glu	Leu	Tyr	Gln	Lys	
		275					280					285	
AGA	GTG	CTG	ACC	ATA	ACC	GGC	ATC	TGC	ATC	GCC	CTC	CTT	896
Arg	Val	Leu	Thr	Ile	Thr	Gly	Ile	Cys	Ile	Ala	Leu	Leu	
				290				295					
GTG	GTC	GGC	ATC	ATG	TGT	GTG	GTG	GCC	TAC	TGC	AAA	ACC	935
Val	Val	Gly	Ile	Met	Cys	Val	Val	Ala	Tyr	Cys	Lys	Thr	
300						305					310		
AAG	AAA	CAG	CGG	AAA	AAG	CTG	CAT	GAC	CGT	CTT	CGG	CAG	974
Lys	Lys	Gln	Arg	Lys	Lys	Leu	His	Asp	Arg	Leu	Arg	Gln	
			315					320					
AGC	CTT	CGG	TCT	GAA	CGA	AAC	AAT	ATG	ATG	AAC	ATT	GCC	1013
Ser	Leu	Arg	Ser	Glu	Arg	Asn	Asn	Met	Met	Asn	Ile	Ala	
325					330					335			

FIG. 1B

3 / 29

AAT	GGG	CCT	CAC	CAT	CCT	AAC	CCA	CCC	CCC	GAG	AAT	GTC	1052
Asn	Gly	Pro	His	His	Pro	Asn	Pro	Pro	Pro	Glu	Asn	Val	
		340					345					350	
CAG	CTG	GTG	AAT	CAA	TAC	GTA	TCT	AAA	AAC	GTC	ATC	TCC	1091
Gln	Leu	Val	Asn	Gln	Tyr	Val	Ser	Lys	Asn	Val	Ile	Ser	
			355						360				
AGT	GAG	CAT	ATT	GTT	GAG	AGA	GAA	GCA	GAG	ACA	TCC	TTT	1130
Ser	Glu	His	Ile	Val	Glu	Arg	Glu	Ala	Glu	Thr	Ser	Phe	
	365					370					375		
TCC	ACC	AGT	CAC	TAT	ACT	TCC	ACA	GCC	CAT	CAC	TCC	ACT	1169
Ser	Thr	Ser	His	Tyr	Thr	Ser	Thr	Ala	His	His	Ser	Thr	
			380					385					
ACT	GTC	ACC	CAG	ACT	CCT	AGC	CAC	AGC	TGG	AGC	AAC	GGA	1208
Thr	Val	Thr	Gln	Thr	Pro	Ser	His	Ser	Trp	Ser	Asn	Gly	
	390				395					400			
CAC	ACT	GAA	AGC	ATC	CTT	TCC	GAA	AGC	CAC	TCT	GTA	ATC	1247
His	Thr	Glu	Ser	Ile	Leu	Ser	Glu	Ser	His	Ser	Val	Ile	
		405					410					415	
GTG	ATG	TCA	TCC	GTA	GAA	AAC	AGT	AGG	CAC	AGC	AGC	CCA	1286
Val	Met	Ser	Ser	Val	Glu	Asn	Ser	Arg	His	Ser	Ser	Pro	
				420				425					
ACT	GGG	GGC	CCA	AGA	GGA	CGT	CTT	AAT	GGC	ACA	GGA	GGC	1325
Thr	Gly	Gly	Pro	Arg	Gly	Arg	Leu	Asn	Gly	Thr	Gly	Gly	
	430					435					440		
CCT	CGT	GAA	TGT	AAC	AGC	TTC	CTC	AGG	CAT	GCC	AGA	GAA	1364
Pro	Arg	Glu	Cys	Asn	Ser	Phe	Leu	Arg	His	Ala	Arg	Glu	
			445					450					
ACC	CCT	GAT	TCC	TAC	CGA	GAC	TCT	CCT	CAT	AGT	GAA	AGG	1403
Thr	Pro	Asp	Ser	Tyr	Arg	Asp	Ser	Pro	His	Ser	Glu	Arg	
					460					465			
TAT	GTG	TCA	GCC	ATG	ACC	ACC	CCG	GCT	CGT	ATG	TCA	CCT	1442
Tyr	Val	Ser	Ala	Met	Thr	Thr	Pro	Ala	Arg	Met	Ser	Pro	
		470					475					480	
GTA	GAT	TTC	CAC	ACG	CCA	AGC	TCC	CCC	AAA	TCG	CCC	CCT	1481
Val	Asp	Phe	His	Thr	Pro	Ser	Ser	Pro	Lys	Ser	Pro	Pro	
				485					490				
TCG	GAA	ATG	TCT	CCA	CCC	GTG	TCC	AGC	ATG	ACG	GTG	TCC	1520
Ser	Glu	Met	Ser	Pro	Pro	Val	Ser	Ser	Met	Thr	Val	Ser	
	495					500					505		

FIG. 1C

4 / 29

ATG	CCT	TCC	ATG	GCG	GTC	AGC	CCC	TTC	ATG	GAA	GAA	GAG	1559
Met	Pro	Ser	Met	Ala	Val	Ser	Pro	Phe	Met	Glu	Glu	Glu	
			510					515					
AGA	CCT	CTA	CTT	CTC	GTG	ACA	CCA	CCA	AGG	CTG	CGG	GAG	1598
Arg	Pro	Leu	Leu	Leu	Val	Thr	Pro	Pro	Arg	Leu	Arg	Glu	
520					525					530			
AAG	AAG	TTT	GAC	CAT	CAC	CCT	CAG	CAG	TTC	AGC	TCC	TTC	1637
Lys	Lys	Phe	Asp	His	His	Pro	Gln	Gln	Phe	Ser	Ser	Phe	
		535					540					545	
CAC	CAC	AAC	CCC	GCG	CAT	GAC	AGT	AAC	AGC	CTC	CCT	GCT	1676
His	His	Asn	Pro	Ala	His	Asp	Ser	Asn	Ser	Leu	Pro	Ala	
				550					555				
AGC	CCC	TTG	AGG	ATA	GTG	GAG	GAT	GAG	GAG	TAT	GAA	ACG	1715
Ser	Pro	Leu	Arg	Ile	Val	Glu	Asp	Glu	Glu	Tyr	Glu	Thr	
	560					565					570		
ACC	CAA	GAG	TAC	GAG	CCA	GCC	CAA	GAG	CCT	GTT	AAG	AAA	1754
Thr	Gln	Glu	Tyr	Glu	Pro	Ala	Gln	Glu	Pro	Val	Lys	Lys	
			575					580					
CTC	GCC	AAT	AGC	CGG	CGG	GCC	AAA	AGA	ACC	AAG	CCC	AAT	1793
Leu	Ala	Asn	Ser	Arg	Arg	Ala	Lys	Arg	Thr	Lys	Pro	Asn	
585					590					595			
GGC	CAC	ATT	GCT	AAC	AGA	TTG	GAA	GTG	GAC	AGC	AAC	ACA	1832
Gly	His	Ile	Ala	Asn	Arg	Leu	Glu	Val	Asp	Ser	Asn	Thr	
		600					605					610	
AGC	TCC	CAG	AGC	AGT	AAC	TCA	GAG	AGT	GAA	ACA	GAA	GAT	1871
Ser	Ser	Gln	Ser	Ser	Asn	Ser	Glu	Ser	Glu	Thr	Glu	Asp	
				615					620				
GAA	AGA	GTA	GGT	GAA	GAT	ACG	CCT	TTC	CTG	GGC	ATA	CAG	1910
Glu	Arg	Val	Gly	Glu	Asp	Thr	Pro	Phe	Leu	Gly	Ile	Gln	
	625					630					635		
AAC	CCC	CTG	GCA	GCC	AGT	CTT	GAG	GCA	ACA	CCT	GCC	TTC	1949
Asn	Pro	Leu	Ala	Ala	Ser	Leu	Glu	Ala	Thr	Pro	Ala	Phe	
			640					645					
CGC	CTG	GCT	GAC	AGC	AGG	ACT	AAC	CCA	GCA	GGC	CGC	TTC	1988
Arg	Leu	Ala	Asp	Ser	Arg	Thr	Asn	Pro	Ala	Gly	Arg	Phe	
650					655					660			
TCG	ACA	CAG	GAA	GAA	ATC	CAG	G						2010
Ser	Thr	Gln	Glu	Glu	Ile	Gln							
		665				669							

FIG. 1D

5 / 29

GG	GAC	AAA	CTT	TTC	CCA	AAC	CCG	ATC	CGA	GCC	CTT	GGA	38
	Asp	Lys	Leu	Phe	Pro	Asn	Pro	Ile	Arg	Ala	Leu	Gly	
	1				5					10			
CCA	AAC	TCG	CCT	GCG	CCG	AGA	GCC	GTC	CGC	GTA	GAG	CGC	77
Pro	Asn	Ser	Pro	Ala	Pro	Arg	Ala	Val	Arg	Val	Glu	Arg	
		15					20					25	
TCC	GTC	TCC	GGC	GAG	ATG	TCC	GAG	CGC	AAA	GAA	GGC	AGA	116
Ser	Val	Ser	Gly	Glu	Met	Ser	Glu	Arg	Lys	Glu	Gly	Arg	
			30						35				
GGC	AAA	GGG	AAG	GGC	AAG	AAG	AAG	GAG	CGA	GGC	TCC	GGC	155
Gly	Lys	Gly	Lys	Gly	Lys	Lys	Lys	Glu	Arg	Gly	Ser	Gly	
	40					45					50		
AAG	AAG	CCG	GAG	TCC	GCG	GCG	GGC	AGC	CAG	AGC	CCA	GCC	194
Lys	Lys	Pro	Glu	Ser	Ala	Ala	Gly	Ser	Gln	Ser	Pro	Ala	
			55					60					
TTG	CCT	CCC	CAA	TTG	AAA	GAG	ATG	AAA	AGC	CAG	GAA	TCG	233
Leu	Pro	Pro	Gln	Leu	Lys	Glu	Met	Lys	Ser	Gln	Glu	Ser	
65					70					75			
GCT	GCA	GGT	TCC	AAA	CTA	GTC	CTT	CGG	TGT	GAA	ACC	AGT	272
Ala	Ala	Gly	Ser	Lys	Leu	Val	Leu	Arg	Cys	Glu	Thr	Ser	
		80					85					90	
TCT	GAA	TAC	TCC	TCT	CTC	AGA	TTC	AAG	TGG	TTC	AAG	AAT	311
Ser	Glu	Tyr	Ser	Ser	Leu	Arg	Phe	Lys	Trp	Phe	Lys	Asn	
				95					100				
GGG	AAT	GAA	TTG	AAT	CGA	AAA	AAC	AAA	CCA	CAA	AAT	ATC	350
Gly	Asn	Glu	Leu	Asn	Arg	Lys	Asn	Lys	Pro	Gln	Asn	Ile	
	105					110					115		
AAG	ATA	CAA	AAA	AAG	CCA	GGG	AAG	TCA	GAA	CTT	CGC	ATT	389
Lys	Ile	Gln	Lys	Lys	Pro	Gly	Lys	Ser	Glu	Leu	Arg	Ile	
			120					125					
AAC	AAA	GCA	TCA	CTG	GCT	GAT	TCT	GGA	GAG	TAT	ATG	TGC	428
Asn	Lys	Ala	Ser	Leu	Ala	Asp	Ser	Gly	Glu	Tyr	Met	Cys	
130					135					140			
AAA	GTG	ATC	AGC	AAA	TTA	GGA	AAT	GAC	AGT	GCC	TCT	GCC	467
Lys	Val	Ile	Ser	Lys	Leu	Gly	Asn	Asp	Ser	Ala	Ser	Ala	
		145					150					155	
AAT	ATC	ACC	ATC	GTG	GAA	TCA	AAC	GAG	ATC	ATC	ACT	GGT	506
Asn	Ile	Thr	Ile	Val	Glu	Ser	Asn	Glu	Ile	Ile	Thr	Gly	
				160					165				

FIG. 2A

6 / 29

ATG	CCA	GCC	TCA	ACT	GAA	GGA	GCA	TAT	GTG	TCT	TCA	GAG	545
Met	Pro	Ala	Ser	Thr	Glu	Gly	Ala	Tyr	Val	Ser	Ser	Glu	
	170					175					180		
TCT	CCC	ATT	AGA	ATA	TCA	GTA	TCC	ACA	GAA	GGA	GCA	AAT	584
Ser	Pro	Ile	Arg	Ile	Ser	Val	Ser	Thr	Glu	Gly	Ala	Asn	
			185					190					
ACT	TCT	TCA	TCT	ACA	TCT	ACA	TCC	ACC	ACT	GGG	ACA	AGC	623
Thr	Ser	Ser	Ser	Thr	Ser	Thr	Ser	Thr	Thr	Gly	Thr	Ser	
195					200					205			
CAT	CTT	GTA	AAA	TGT	GCG	GAG	AAG	GAG	AAA	ACT	TTC	TGT	662
His	Leu	Val	Lys	Cys	Ala	Glu	Lys	Glu	Lys	Thr	Phe	Cys	
		210					215					220	
GTG	AAT	GGA	GGG	GAG	TGC	TTC	ATG	GTG	AAA	GAC	CTT	TCA	701
Val	Asn	Gly	Gly	Glu	Cys	Phe	Met	Val	Lys	Asp	Leu	Ser	
				225					230				
AAC	CCC	TCG	AGA	TAC	TTG	TGC	AAG	TGC	CCA	AAT	GAG	TTT	740
Asn	Pro	Ser	Arg	Tyr	Leu	Cys	Lys	Cys	Pro	Asn	Glu	Phe	
	235					240					245		
ACT	GGT	GAT	CGC	TGC	CAA	AAC	TAC	GTA	ATG	GCC	AGC	TTC	779
Thr	Gly	Asp	Arg	Cys	Gln	Asn	Tyr	Val	Met	Ala	Ser	Phe	
			250					255					
TAC	AAG	CAT	CTT	GGG	ATT	GAA	TTT	ATG	GAG	GCG	GAG	GAG	818
Tyr	Lys	His	Leu	Gly	Ile	Glu	Phe	Met	Glu	Ala	Glu	Glu	
260					265					270			
CTG	TAC	CAG	AAG	AGA	GTG	CTG	ACC	ATA	ACC	GGC	ATC	TGC	857
Leu	Tyr	Gln	Lys	Arg	Val	Leu	Thr	Ile	Thr	Gly	Ile	Cys	
		275					280					285	
ATC	GCC	CTC	CTT	GTG	GTC	GGC	ATC	ATG	TGT	GTG	GTG	GCC	896
Ile	Ala	Leu	Leu	Val	Val	Gly	Ile	Met	Cys	Val	Val	Ala	
				290					295				
TAC	TGC	AAA	ACC	AAG	AAA	CAG	CGG	AAA	AAG	CTG	CAT	GAC	935
Tyr	Cys	Lys	Thr	Lys	Lys	Gln	Arg	Lys	Lys	Leu	His	Asp	
	300					305					310		
CGT	CTT	CGG	CAG	AGC	CTT	CGG	TCT	GAA	CGA	AAC	AAT	ATG	974
Arg	Leu	Arg	Gln	Ser	Leu	Arg	Ser	Glu	Arg	Asn	Asn	Met	
			315					320					
ATG	AAC	ATT	GCC	AAT	GGG	CCT	CAC	CAT	CCT	AAC	CCA	CCC	1013
Met	Asn	Ile	Ala	Asn	Gly	Pro	His	His	Pro	Asn	Pro	Pro	
325					330					335			

FIG. 2B

7 / 29

CCC	GAG	AAT	GTC	CAG	CTG	GTG	AAT	CAA	TAC	GTA	TCT	AAA	1052
Pro	Glu	Asn	Val	Gln	Leu	Val	Asn	Gln	Tyr	Val	Ser	Lys	
		340					345					350	
AAC	GTC	ATC	TCC	AGT	GAG	CAT	ATT	GTT	GAG	AGA	GAA	GCA	1091
Asn	Val	Ile	Ser	Ser	Glu	His	Ile	Val	Glu	Arg	Glu	Ala	
				355					360				
GAG	ACA	TCC	TTT	TCC	ACC	AGT	CAC	TAT	ACT	TCC	ACA	GCC	1130
Glu	Thr	Ser	Phe	Ser	Thr	Ser	His	Tyr	Thr	Ser	Thr	Ala	
	365					370					375		
CAT	CAC	TCC	ACT	ACT	GTC	ACC	CAG	ACT	CCT	AGC	CAC	AGC	1169
His	His	Ser	Thr	Thr	Val	Thr	Gln	Thr	Pro	Ser	His	Ser	
			380					385					
TGG	AGC	AAC	GGA	CAC	ACT	GAA	AGC	ATC	CTT	TCC	GAA	AGC	1208
Trp	Ser	Asn	Gly	His	Thr	Glu	Ser	Ile	Leu	Ser	Glu	Ser	
	390					395				400			
CAC	TCT	GTA	ATC	GTG	ATG	TCA	TCC	GTA	GAA	AAC	AGT	AGG	1247
His	Ser	Val	Ile	Val	Met	Ser	Ser	Val	Glu	Asn	Ser	Arg	
		405					410					415	
CAC	AGC	AGC	CCA	ACT	GGG	GGC	CCA	AGA	GGA	CGT	CTT	AAT	1286
His	Ser	Ser	Pro	Thr	Gly	Gly	Pro	Arg	Gly	Arg	Leu	Asn	
				420					425				
GGC	ACA	GGA	GGC	CCT	CGT	GAA	TGT	AAC	AGC	TTC	CTC	AGG	1325
Gly	Thr	Gly	Gly	Pro	Arg	Glu	Cys	Asn	Ser	Phe	Leu	Arg	
	430					435					440		
CAT	GCC	AGA	GAA	ACC	CCT	GAT	TCC	TAC	CGA	GAC	TCT	CCT	1364
His	Ala	Arg	Glu	Thr	Pro	Asp	Ser	Tyr	Arg	Asp	Ser	Pro	
			445					450					
CAT	AGT	GAA	AGG	TAT	GTG	TCA	GCC	ATG	ACC	ACC	CCG	GCT	1403
His	Ser	Glu	Arg	Tyr	Val	Ser	Ala	Met	Thr	Thr	Pro	Ala	
	455				460					465			
CGT	ATG	TCA	CCT	GTA	GAT	TTC	CAC	ACG	CCA	AGC	TCC	CCC	1442
Arg	Met	Ser	Pro	Val	Asp	Phe	His	Thr	Pro	Ser	Ser	Pro	
			470				475					480	
AAA	TCG	CCC	CCT	TCG	GAA	ATG	TCT	CCA	CCC	GTG	TCC	AGC	1481
Lys	Ser	Pro	Pro	Ser	Glu	Met	Ser	Pro	Pro	Val	Ser	Ser	
				485					490				
ATG	ACG	GTG	TCC	ATG	CCT	TCC	ATG	GCG	GTG	AGC	CCC	TTC	1520
Met	Thr	Val	Ser	Met	Pro	Ser	Met	Ala	Val	Ser	Pro	Phe	
	495					500					505		

FIG. 2C

8 / 29

ATG	GAA	GAA	GAG	AGA	CCT	CTA	CTT	CTC	GTG	ACA	CCA	CCA	1559
Met	Glu	Glu	Glu	Arg	Pro	Leu	Leu	Leu	Val	Thr	Pro	Pro	
			510					515					
AGG	CTG	CGG	GAG	AAG	AAG	TTT	GAC	CAT	CAC	CCT	CAG	CAG	1598
Arg	Leu	Arg	Glu	Lys	Lys	Phe	Asp	His	His	Pro	Gln	Gln	
520					525					530			
TTC	AGC	TCC	TTC	CAC	CAC	AAC	CCC	GCG	CAT	GAC	AGT	AAC	1637
Phe	Ser	Ser	Phe	His	His	Asn	Pro	Ala	His	Asp	Ser	Asn	
		535					540					545	
AGC	CTC	CCT	GCT	AGC	CCC	TTG	AGG	ATA	GTG	GAG	GAT	GAG	1676
Ser	Leu	Pro	Ala	Ser	Pro	Leu	Arg	Ile	Val	Glu	Asp	Glu	
			550						555				
GAG	TAT	GAA	ACG	ACC	CAA	GAG	TAC	GAG	CCA	GCC	CAA	GAG	1715
Glu	Tyr	Glu	Thr	Thr	Gln	Glu	Tyr	Glu	Pro	Ala	Gln	Glu	
	560					565					570		
CCT	GTT	AAG	AAA	CTC	GCC	AAT	AGC	CGG	CGG	GCC	AAA	AGA	1754
Pro	Val	Lys	Lys	Leu	Ala	Asn	Ser	Arg	Arg	Ala	Lys	Arg	
			575					580					
ACC	AAG	CCC	AAT	GGC	CAC	ATT	GCT	AAC	AGA	TTG	GAA	GTG	1793
Thr	Lys	Pro	Asn	Gly	His	Ile	Ala	Asn	Arg	Leu	Glu	Val	
585					590					595			
GAC	AGC	AAC	ACA	AGC	TCC	CAG	AGC	AGT	AAC	TCA	GAG	AGT	1832
Asp	Ser	Asn	Thr	Ser	Ser	Gln	Ser	Ser	Asn	Ser	Glu	Ser	
		600					605					610	

FIG. 2D

9 / 29

GAA	ACA	GAA	GAT	GAA	AGA	GTA	GGT	GAA	GAT	ACG	CCT	TTC	1871
Glu	Thr	Glu	Asp	Glu	Arg	Val	Gly	Glu	Asp	Thr	Pro	Phe	
				615					620				
CTG	GGC	ATA	CAG	AAC	CCC	CTG	GCA	GCC	AGT	CTT	GAG	GCA	1910
Leu	Gly	Ile	Gln	Asn	Pro	Leu	Ala	Ala	Ser	Leu	Glu	Ala	
	625					630					635		
ACA	CCT	GCC	TTC	CGC	CTG	GCT	GAC	AGC	AGG	ACT	AAC	CCA	1949
Thr	Pro	Ala	Phe	Arg	Leu	Ala	Asp	Ser	Arg	Thr	Asn	Pro	
			640					645					
GCA	GGC	CGC	TTC	TCG	ACA	CAG	GAA	GAA	ATC	CAG	GCC	AGG	1988
Ala	Gly	Arg	Phe	Ser	Thr	Gln	Glu	Glu	Ile	Gln	Ala	Arg	
650					655					660			
CTG	TCT	AGT	GTA	ATT	GCT	AAC	CAA	GAC	CCT	ATT	GCT	GTA	TA 2029
Leu	Ser	Ser	Val	Ile	Ala	Asn	Gln	Asp	Pro	Ile	Ala	Val	
		665					670					675	
A AACCTAAATA AACACATAGA TTCACCTGTA AAACCTTTATT													2070
TTATATAATA AAGTATTCCA CCTTAAATTA AACCAATTTAT TTTATTTTAG													2120
CAGTTCTGCA AATAGAAAAC AGGAAAAAAA CTTTTATAAA TTAAATATAT													2170
GTATGTAAAA ATGAAAAAAA AAAAAAAA													2199

FIG. 2E

10 / 29

GTGGCTGCGG GGCAATTGAA AAAGAGCCGG CGAGGAGTTC CCCGAAACTT 50
 GTTGGAAGTC CGGGCTCGCG CGGAGGCCAG GAGCTGAGCG GCGGCGGCTG 100
 COGGACGATG GGAGCGTGAG CAGGACGGTG ATAACCTCTC CCCGATCGGG 150
 TTGCGAGGGC GCCGGGCAGA GGCCAGGACG CGAGCCGCCA GCGGCGGGAC 200
 CCATCGACGA CTTCCCGGGG CGACAGGAGC AGCCCCGAGA GCCAGGGCGA 250
 GCGCCCGTTC CAGGTGGCCG GACCGCCCGC CGCGTCCGCG CCGCGCTCCC 300
 TGCAGGCAAC GGGAGACGCC CCCGCGCAGC GCGAGCGCCT CAGCGCGGCC 350
 GCTCGCTCTC CCCATCGAGG GACAACTTT TCCCAAACCC GATCCGAGCC 400
 CTTGGACCAA ACTCGCCTGC GCCGAGAGCC GTCCGCGTAG AGCGCTCCGT 450
 CTCCGGCGAG ATG TCC GAG CGC AAA GAA GGC AGA GGC AAA 490
 Met Ser Glu Arg Lys Glu Gly Arg Gly Lys
 1 5 10
 GGG AAG GGC AAG AAG AAG GAG CGA GGC TCC GGC AAG AAG 529
 Gly Lys Gly Lys Lys Lys Glu Arg Gly Ser Gly Lys Lys
 15 20
 CCG GAG TCC GCG GCG GGC AGC CAG AGC CCA GCC TTG CCT 568
 Pro Glu Ser Ala Ala Gly Ser Gln Ser Pro Ala Leu Pro
 25 30 35
 CCC CAA TTG AAA GAG ATG AAA AGC CAG GAA TCG GCT GCA 607
 Pro Gln Leu Lys Glu Met Lys Ser Gln Glu Ser Ala Ala
 40 45
 GGT TCC AAA CTA GTC CTT CGG TGT GAA ACC AGT TCT GAA 646
 Gly Ser Lys Leu Val Leu Arg Cys Glu Thr Ser Ser Glu
 50 55 60
 TAC TCC TCT CTC AGA TTC AAG TGG TTC AAG AAT GGG AAT 685
 Tyr Ser Ser Leu Arg Ph Lys Trp Phe Lys Asn Gly Asn
 65 70 75
 GAA TTG AAT CGA AAA AAC AAA CCA CAA AAT ATC AAG ATA 724
 Glu Leu Asn Arg Lys Asn Lys Pro Gln Asn Ile Lys Ile
 80 85

FIG. 3A

11 / 29

CAA	AAA	AAG	CCA	GGG	AAG	TCA	GAA	CTT	CGC	ATT	AAC	AAA	763
Gln	Ly	Lys	Pro	Gly	Lys	S r	Glu	Leu	Arg	Il	Asn	Lys	
	90					95					100		
GCA	TCA	CTG	GCT	GAT	TCT	GGA	GAG	TAT	ATG	TGC	AAA	GTG	802
Ala	Ser	Leu	Ala	Asp	Ser	Gly	Glu	Tyr	Met	Cys	Lys	Val	
			105					110					
ATC	AGC	AAA	TTA	GGA	AAT	GAC	AGT	GCC	TCT	GCC	AAT	ATC	841
Ile	Ser	Lys	Leu	Gly	Asn	Asp	Ser	Ala	Ser	Ala	Asn	Ile	
115					120					125			
ACC	ATC	GTG	GAA	TCA	AAC	GAG	ATC	ATC	ACT	GGT	ATG	CCA	880
Thr	Ile	Val	Glu	Ser	Asn	Glu	Ile	Ile	Thr	Gly	Met	Pro	
		130					135					140	
GCC	TCA	ACT	GAA	GGA	GCA	TAT	GTG	TCT	TCA	GAG	TCT	CCC	919
Ala	Ser	Thr	Glu	Gly	Ala	Tyr	Val	Ser	Ser	Glu	Ser	Pro	
				145					150				
ATT	AGA	ATA	TCA	GTA	TCC	ACA	GAA	GGA	GCA	AAT	ACT	TCT	958
Ile	Arg	Ile	Ser	Val	Ser	Thr	Glu	Gly	Ala	Asn	Thr	Ser	
	155					160					165		
TCA	TCT	ACA	TCT	ACA	TCC	ACC	ACT	GGG	ACA	AGC	CAT	CTT	997
Ser	Ser	Thr	Ser	Thr	Ser	Thr	Thr	Gly	Thr	Ser	His	Leu	
			170					175					
GTA	AAA	TGT	GCG	GAG	AAG	GAG	AAA	ACT	TTC	TGT	GTG	AAT	1036
Val	Lys	Cys	Ala	Glu	Lys	Glu	Lys	Thr	Phe	Cys	Val	Asn	
180					185					190			
GGA	GGG	GAG	TGC	TTC	ATG	GTG	AAA	GAC	CTT	TCA	AAC	CCC	1075
Gly	Gly	Glu	Cys	Phe	Met	Val	Lys	Asp	Leu	Ser	Asn	Pro	
		195					200					205	
TCG	AGA	TAC	TTG	TGC	AAG	TGC	CCA	AAT	GAG	TTT	ACT	GGT	1114
Ser	Arg	Tyr	Leu	Cys	Lys	Cys	Pro	Asn	Glu	Phe	Thr	Gly	
				210					215				
GAT	CGC	TGC	CAA	AAC	TAC	GTA	ATG	GCC	AGC	TTC	TAC	AAG	1153
Asp	Arg	Cys	Gln	Asn	Tyr	Val	Met	Ala	Ser	Phe	Tyr	Lys	
	220					225					230		
GCG	GAG	GAG	CTG	TAC	CAG	AAG	AGA	GTG	CTG	ACC	ATA	ACC	1192
Ala	Glu	Glu	Leu	Tyr	Gln	Lys	Arg	Val	Leu	Thr	Ile	Thr	
			235					240					
GGC	ATC	TGC	ATC	GCC	CTC	CTT	GTG	GTC	GGC	ATC	ATG	TGT	1231
Gly	Ile	Cys	Ile	Ala	Leu	Leu	Val	Val	Gly	Ile	Met	Cys	
245					250					255			
GTG	GTG	GCC	TAC	TGC	AAA	ACC	AAG	AAA	CAG	CGG	AAA	AAG	1270
Val	Val	Ala	Tyr	Cys	Lys	Thr	Lys	Lys	Gln	Arg	Lys	Lys	
		260					265					270	

FIG. 3B

12/29

CTG	CAT	GAC	CGT	CTT	CGG	CAG	AGC	CTT	CGG	TCT	GAA	CGA	1309
Leu	His	Asp	Arg	Leu	Arg	Gln	Ser	Leu	Arg	Ser	Glu	Arg	
				275					280				

AAC	AAT	ATG	ATG	AAC	ATT	GCC	AAT	GGG	CCT	CAC	CAT	CCT	1348
Asn	Asn	Met	Met	Asn	Ile	Ala	Asn	Gly	Pro	His	His	Pro	
	285					290					295		

AAC	CCA	CCC	CCC	GAG	AAT	GTC	CAG	CTG	GTG	AAT	CAA	TAC	1387
Asn	Pro	Pro	Pro	Glu	Asn	Val	Gln	Leu	Val	Asn	Gln	Tyr	
			300					305					

GTA	TCT	AAA	AAC	GTC	ATC	TCC	AGT	GAG	CAT	ATT	GTT	GAG	1426
Val	Ser	Lys	Asn	Val	Ile	Ser	Ser	Glu	His	Ile	Val	Glu	
310					315					320			

AGA	GAA	GCA	GAG	ACA	TCC	TTT	TCC	ACC	AGT	CAC	TAT	ACT	1465
Arg	Glu	Ala	Glu	Thr	Ser	Phe	Ser	Thr	Ser	His	Tyr	Thr	
		325					330					335	

TCC	ACA	GCC	CAT	CAC	TCC	ACT	ACT	GTC	ACC	CAG	ACT	CCT	1504
Ser	Thr	Ala	His	His	Ser	Thr	Thr	Val	Thr	Gln	Thr	Pro	
				340						345			

AGC	CAC	AGC	TGG	AGC	AAC	GGA	CAC	ACT	GAA	AGC	ATC	CTT	1543
Ser	His	Ser	Trp	Ser	Asn	Gly	His	Thr	Glu	Ser	Ile	Leu	
	350					355					360		

TCC	GAA	AGC	CAC	TCT	GTA	ATC	GTG	ATG	TCA	TCC	GTA	GAA	1582
Ser	Glu	Ser	His	Ser	Val	Ile	Val	Met	Ser	Ser	Val	Glu	
			365					370					

AAC	AGT	AGG	CAC	AGC	AGC	CCA	ACT	GGG	GGC	CCA	AGA	GGA	1621
Asn	Ser	Arg	His	Ser	Ser	Pro	Thr	Gly	Gly	Pro	Arg	Gly	
375					380					385			

CGT	CTT	AAT	GGC	ACA	GGA	GGC	CCT	CGT	GAA	TGT	AAC	AGC	1660
Arg	Leu	Asn	Gly	Thr	Gly	Gly	Pro	Arg	Glu	Cys	Asn	Ser	
		390					395					400	

TTC	CTC	AGG	CAT	GCC	AGA	GAA	ACC	CCT	GAT	TCC	TAC	CGA	1699
Phe	Leu	Arg	His	Ala	Arg	Glu	Thr	Pro	Asp	Ser	Tyr	Arg	
				405					410				

GAC	TCT	CCT	CAT	AGT	GAA	AGG	TAT	GTG	TCA	GCC	ATG	ACC	1738
Asp	Ser	Pro	His	Ser	Glu	Arg	Tyr	Val	Ser	Ala	Met	Thr	
	415					420					425		

ACC	CCG	GCT	CGT	ATG	TCA	CCT	GTA	GAT	TTC	CAC	ACG	CCA	1777
Thr	Pr	Ala	Arg	Met	Ser	Pro	Val	Asp	Phe	His	Thr	Pr	
			430					435					

AGC	TCC	CCC	AAA	TCG	CCC	CCT	TCG	GAA	ATG	TCT	CCA	CCC	1816
Ser	Ser	Pr	Lys	Ser	Pro	Pro	Ser	Glu	Met	Ser	Pro	Pro	
440					445					450			

FIG._3C

13/29

GTG	TCC	AGC	ATG	ACG	GTG	TCC	AAG	CCT	TCC	ATG	GCG	GTG	1855
Val	Ser	Ser	Met	Thr	Val	Ser	Lys	Pro	Ser	Met	Ala	Val	
		455					460					465	
AGC	CCC	TTC	ATG	GAA	GAA	GAG	AGA	CCT	CTA	CTT	CTC	GTG	1894
Ser	Pro	Phe	Met	Glu	Glu	Glu	Arg	Pro	Leu	Leu	Leu	Val	
				470					475				
ACA	CCA	CCA	AGG	CTG	CGG	GAG	AAG	AAG	TTT	GAC	CAT	CAC	1933
Thr	Pro	Pro	Arg	Leu	Arg	Glu	Lys	Lys	Phe	Asp	His	His	
		480				485					490		
CCT	CAG	CAG	TTC	AGC	TCC	TTC	CAC	CAC	AAC	CCC	GCG	CAT	1972
Pro	Gln	Gln	Phe	Ser	Ser	Phe	His	His	Asn	Pro	Ala	His	
			495					500					
GAC	AGT	AAC	AGC	CTC	CCT	GCT	AGC	CCC	TTG	AGG	ATA	GTG	2011
Asp	Ser	Asn	Ser	Leu	Pro	Ala	Ser	Pro	Leu	Arg	Ile	Val	
505					510					515			
GAG	GAT	GAG	GAG	TAT	GAA	ACG	ACC	CAA	GAG	TAC	GAG	CCA	2050
Glu	Asp	Glu	Glu	Tyr	Glu	Thr	Thr	Gln	Glu	Tyr	Glu	Pro	
		520					525					530	
GCC	CAA	GAG	CCT	GTT	AAG	AAA	CTC	GCC	AAT	AGC	CGG	CGG	2089
Ala	Gln	Glu	Pro	Val	Lys	Lys	Leu	Ala	Asn	Ser	Arg	Arg	
				535					540				
GCC	AAA	AGA	ACC	AAG	CCC	AAT	GGC	CAC	ATT	GCT	AAC	AGA	2128
Ala	Lys	Arg	Thr	Lys	Pro	Asn	Gly	His	Ile	Ala	Asn	Arg-	
	545					550					555		
TTG	GAA	GTG	GAC	AGC	AAC	ACA	AGC	TCC	CAG	AGC	AGT	AAC	2167
Leu	Glu	Val	Asp	Ser	Asn	Thr	Ser	Ser	Gln	Ser	Ser	Asn	
			560					565					
TCA	GAG	AGT	GAA	ACA	GAA	GAT	GAA	AGA	GTA	GGT	GAA	GAT	2206
Ser	Glu	Ser	Glu	Thr	Glu	Asp	Glu	Arg	Val	Gly	Glu	Asp	
570					575					580			
ACG	CCT	TTC	CTG	GGC	ATA	CAG	AAC	CCC	CTG	GCA	GCC	AGT	2245
Thr	Pro	Phe	Leu	Gly	Ile	Gln	Asn	Pro	Leu	Ala	Ala	Ser	
		585					590					595	
CTT	GAG	GCA	ACA	CCT	GCC	TTC	CGC	CTG	GCT	GAC	AGC	AGG	2284
Leu	Glu	Ala	Thr	Pro	Ala	Phe	Arg	Leu	Ala	Asp	Ser	Arg	
				600					605				
ACT	AAC	CCA	GCA	GGC	CGC	TTC	TCG	ACA	CAG	GAA	GAA	ATC	2323
Thr	Asn	Pro	Ala	Gly	Arg	Ph	Ser	Thr	Gln	Glu	Glu	Ile	
	610					615					620		
CAG	GCC	AGG	CTG	TCT	AGT	GTA	ATT	GCT	AAC	CAA	GAC	CCT	2362
Gln	Ala	Arg	Leu	Ser	Ser	Val	Ile	Ala	Asn	Gln	Asp	Pro	
			625					630					

FIG. 3D

14 / 29

ATT GCT GTA TAAAACCTA AATAACACA TAGATTCACC TGTAAAACTT 2410
Ile Ala Val
635 637

TATTTTATAT AATAAAGTAT TCCACCTTAA ATTAAACAAT TTATTTTATT 2460

TTAGCAGTTC TGCAAATAAA AAAAAAAAAA 2490

FIG._3E

GCGCCTGCOCT CCAACCTGCG GGCGGGAGGT GGGTGGCTGC GGGGCAATTG 50

AAAAAGAGCC GGCGAGGAGT TCCCCGAAAC TTGTTGGAAC TCCGGGCTCG 100

CGCGGAGGCC AGGAGCTGAG CGGCGGCGGC TGCCGGACGA TGGGAGCGTG 150

AGCAGGACGG TGATAACCTC TCCCCGATCG GGTTCGAGG GCGCCGGGCA 200

GAGGCCAGGA CGCGAGCCGC CAGCGGCGGG ACCCATCGAC GACTTCCCGG 250

GGCGACAGGA GCAGCCCCGA GAGCCAGGGC GAGCGCCCGT TCCAGGTGGC 300

CGGACCGCCC GCCGCGTCCG CGCCGCGCTC CCTGCAGGCA ACGGGAGACG 350

CCCCCGCGCA GCGCGAGCGC CTCAGCGCGG CCGCTCGCTC TCCCCATCGA 400

GGGACAAACT TTTCCCAAAC CCGATCCGAG CCCTTGGACC AAACCTCGCCT 450

GCGCCGAGAG CCGTCCGCGT AGAGCGCTCC GTCTCCGGCG AG ATG 495
Met
1

TCC GAG CGC AAA GAA GGC AGA GGC AAA GGG AAG GGC AAG 534
Ser Glu Arg Lys Glu Gly Arg Gly Lys Gly Lys Gly Lys
5 10

AAG AAG GAG CGA GGC TCC GGC AAG AAG CCG GAG TCC GCG 573
Lys Lys Glu Arg Gly Ser Gly Lys Lys Pro Glu Ser Ala
15 20 25

GCG GGC AGC CAG AGC CCA GCC TTG CCT CCC CAA TTG AAA 612
Ala Gly Ser Gln Ser Pro Ala Leu Pro Pro Gln Leu Lys
30 35 40

GAG ATG AAA AGC CAG GAA TCG GCT GCA GGT TCC AAA CTA 651
Glu Met Lys Ser Gln Glu Ser Ala Ala Gly Ser Lys Leu
45 50

GTC CTT CGG TGT GAA ACC AGT TCT GAA TAC TCC TCT CTC 690
Val Leu Arg Cys Glu Thr Ser Ser Glu Tyr Ser Ser Leu
55 60 65

AGA TTC AAG TGG TTC AAG AAT GGG AAT GAA TTG AAT CGA 729
Arg Phe Lys Trp Phe Lys Asn Gly Asn Glu Leu Asn Arg
70 75

FIG. 4A

16 / 29

AAA AAC AAA CCA CAA AAT ATC AAG ATA CAA AAA AAG CCA 768
 Lys Asn Lys Pro Gln Asn Ile Lys Ile Gln Lys Lys Pro
 80 85 90

GGA AAG TCA GAA CTT CGC ATT AAC AAA GCA TCA CTG GCT 807
 Gly Lys Ser Glu Leu Arg Ile Asn Lys Ala Ser Leu Ala
 95 100 105

GAT TCT GGA GAG TAT ATG TGC AAA GTG ATC AGC AAA TTA 846
 Asp Ser Gly Glu Tyr Met Cys Lys Val Ile Ser Lys Leu
 110 115

GGA AAT GAC AGT GCC TCT GCC AAT ATC ACC ATC GTG GAA 885
 Gly Asn Asp Ser Ala Ser Ala Asn Ile Thr Ile Val Glu
 120 125 130

TCA AAC GAG ATC ATC ACT GGT ATG CCA GCC TCA ACT GAA 924
 Ser Asn Glu Ile Ile Thr Gly Met Pro Ala Ser Thr Glu
 135 140

GGA GCA TAT GTG TCT TCA GAG TCT CCC ATT AGA ATA TCA 963
 Gly Ala Tyr Val Ser Ser Glu Ser Pro Ile Arg Ile Ser
 145 150 155

GTA TCC ACA GAA GGA GCA AAT ACT TCT TCA TCT ACA TCT 1002
 Val Ser Thr Glu Gly Ala Asn Thr Ser Ser Ser Thr Ser
 160 165 170

ACA TCC ACC ACT GGG ACA AGC CAT CTT GTA AAA TGT GCG 1041
 Thr Ser Thr Thr Gly Thr Ser His Leu Val Lys Cys Ala
 175 180

GAG AAG GAG AAA ACT TTC TGT GTG AAT GGA GGG GAG TGC 1080
 Glu Lys Glu Lys Thr Phe Cys Val Asn Gly Gly Glu Cys
 185 190 195

TTC ATG GTG AAA GAC CTT TCA AAC CCC TCG AGA TAC TTG 1119
 Phe Met Val Lys Asp Leu Ser Asn Pro Ser Arg Tyr Leu
 200 205

TGC AAG TGC CCA AAT GAG TTT ACT GGT GAT CGC TGC CAA 1158
 Cys Lys Cys Pro Asn Glu Phe Thr Gly Asp Arg Cys Gln
 210 215 220

AAC TAC GTA ATG GCC AGC TTC TAC AGT ACG TCC ACT CCC 1197
 Asn Tyr Val Met Ala Ser Phe Tyr Ser Thr Ser Thr Pro
 225 230 235

TTT CTG TCT CTG CCT GAA TAGGA GCATGCTCAG TTGGTGCTGC 1240
 Phe Leu Ser Leu Pro Glu
 240 241

TTTCTTGTTG CTGCATCTCC CCTCAGATTC CACCTAGAGC TAGATGTGTC 1290

FIG. 4B

17 / 29

TTACCAGATC TAATATTGAC TGCCTCTGCC TGTCGCATGA GAACATTAAC 1340
AAAAGCAATT GTATTACTTC CTCTGTTCGC GACTAGTTGG CTCTGAGATA 1390
CTAATAGGTG TGTGAGGCTC CGGATGTTTC TGGAATTGAT ATTGAATGAT 1440
GTGATACAAA TTGATAGTCA ATATCAAGCA GTGAAATATG ATAATAAAGG 1490
CATTTCAAAG TCTCACTTTT ATTGATAAAA TAAAAATCAT TCTACTGAAC 1540
AGTCCATCTT CTTTATACAA TGACCACATC CTGAAAAGGG TGTGCTAAG 1590
CTGTAACCGA TATGCACTTG AAATGATGGT AAGTTAATTT TGATTCAGAA 1640
TGTGTTATTT GTCACAAATA AACATAATAA AAGGAGTTCA GATGTTTTTC 1690
TTCATTAACC AAAAAAAAAA AAAAA 1715

FIG. 4C

18 / 29

GAGGCGCCTG CCTCCAACCT GCGGGCGGGA GGTGGGTGGC TCGGGGGCAA 50
 TTGAAAAGA GCGGGCGAGG AGTTCCCCGA AACTTGTTGG AACTCCGGGC 100
 TCGCGCGGAG GCCAGGAGCT GAGCGGCGGC GGCTGCCGGA CGATGGGAGC 150
 GTGAGCAGGA CGGTGATAAC CTCTCCCCGA TCGGGTTGCG AGGGCGCCGG 200
 GCAGAGGCCA GGACGCGAGC CGCCAGCGGC GGGACCCATC GACGACTTCC 250
 CGGGGCGACA GGAGCAGCCC CGAGAGCCAG GGCAGCGGCC CGTTCCAGGT 300
 GGCCGGACCG CCCGCCGCGT CCGCGCCGCG CTCCCTGCAG GCAACGGGAG 350
 ACGCCCCCGC GCAGCGCGAG CGCCTCAGCG CGGCCGCTCG CTCTCCCCAT 400
 CGAGGGACAA ACTTTTCCCA AACCCGATCC GAGCCCTTGG ACCAACTCG 450
 CCTGCGCCGA GAGCCGTCCG CGTAGAGCGC TCCGTCTCCG GCGAG AT 497
 Met
 - 1
 G TCC GAG CGC AAA GAA GGC AGA GGC AAA GGG AAG GGC AAG 537
 Ser Glu Arg Lys Glu Gly Arg Gly Lys Gly Lys Gly Lys
 5 10
 AAG AAG GAG CGA GGC TCC GGC AAG AAG CCG GAG TCC GCG 576
 Lys Lys Glu Arg Gly Ser Gly Lys Lys Pro Glu Ser Ala
 15 20 25
 GCG GGC AGC CAG AGC CCA GCC TTG CCT CCC CAA TTG AAA 615
 Ala Gly Ser Gln Ser Pro Ala Leu Pro Pro Gln Leu Lys
 30 35 40
 GAG ATG AAA AGC CAG GAA TCG GCT GCA GGT TCC AAA CTA 654
 Glu Met Lys Ser Gln Glu Ser Ala Ala Gly Ser Lys Leu
 45 50
 GTC CTT CGG TGT GAA ACC AGT TCT GAA TAC TCC TCT CTC 693
 Val Leu Arg Cys Glu Thr Ser Ser Glu Tyr Ser Ser Leu
 55 60 65
 AGA TTC AAG TGG TTC AAG AAT GGG AAT GAA TTG AAT CGA 732
 Arg Phe Lys Trp Phe Lys Asn Gly Asn Glu Leu Asn Arg
 70 75

FIG. 5A

19 / 29

AAA	AAC	AAA	CCA	CAA	AAT	ATC	AAG	ATA	CAA	AAA	AAG	CCA	771
Lys	Asn	Lys	Pro	Gln	Asn	Ile	Lys	Ile	Gln	Lys	Lys	Pro	
80					85					90			
GGG	AAG	TCA	GAA	CTT	CGC	ATT	AAC	AAA	GCA	TCA	CTG	GCT	810
Gly	Lys	Ser	Glu	Leu	Arg	Ile	Asn	Lys	Ala	Ser	Leu	Ala	
	95						100					105	
GAT	TCT	GGA	GAG	TAT	ATG	TGC	AAA	GTG	ATC	AGC	AAA	TTA	849
Asp	Ser	Gly	Glu	Tyr	Met	Cys	Lys	Val	Ile	Ser	Lys	Leu	
				110					115				
GGA	AAT	GAC	AGT	GCC	TCT	GCC	AAT	ATC	ACC	ATC	GTG	GAA	888
Gly	Asn	Asp	Ser	Ala	Ser	Ala	Asn	Ile	Thr	Ile	Val	Glu	
	120					125					130		
TCA	AAC	GAG	ATC	ATC	ACT	GGT	ATG	CCA	GCC	TCA	ACT	GAA	927
Ser	Asn	Glu	Ile	Ile	Thr	Gly	Met	Pro	Ala	Ser	Thr	Glu	
			135					140					
GGA	GCA	TAT	GTG	TCT	TCA	GAG	TCT	CCC	ATT	AGA	ATA	TCA	966
Gly	Ala	Tyr	Val	Ser	Ser	Glu	Ser	Pro	Ile	Arg	Ile	Ser	
145					150					155			
GTA	TCC	ACA	GAA	GGA	GCA	AAT	ACT	TCT	TCA	TCT	ACA	TCT	1005
Val	Ser	Thr	Glu	Gly	Ala	Asn	Thr	Ser	Ser	Ser	Thr	Ser	
		160					165					170	
ACA	TCC	ACC	ACT	GGG	ACA	AGC	CAT	CTT	GTA	AAA	TGT	GCG	1044
Thr	Ser	Thr	Thr	Gly	Thr	Ser	His	Leu	Val	Lys	Cys	Ala	
				175					180				
GAG	AAG	GAG	AAA	ACT	TTC	TGT	GTG	AAT	GGA	GGG	GAG	TGC	1083
Glu	Lys	Glu	Lys	Thr	Phe	Cys	Val	Asn	Gly	Gly	Glu	Cys	
	185					190					195		
TTC	ATG	GTG	AAA	GAC	CTT	TCA	AAC	CCC	TCG	AGA	TAC	TTG	1122
Phe	Met	Val	Lys	Asp	Leu	Ser	Asn	Pro	Ser	Arg	Tyr	Leu	
			200					205					
TGC	AAG	TGC	CCA	AAT	GAG	TTT	ACT	GGT	GAT	CGC	TGC	CAA	1161
Cys	Lys	Cys	Pro	Asn	Glu	Phe	Thr	Gly	Asp	Arg	Cys	Gln	
210					215					220			
AAC	TAC	GTA	ATG	GCC	AGC	TTC	TAC	AAG	GCG	GAG	GAG	CTG	1200
Asn	Tyr	Val	Met	Ala	Ser	Phe	Tyr	Lys	Ala	Glu	Glu	Leu	
		225					230					235	
TAC	CAG	AAG	AGA	GTG	CTG	ACC	ATA	ACC	GGC	ATC	TGC	ATC	1239
Tyr	Gln	Lys	Arg	Val	Leu	Thr	Ile	Thr	Gly	Ile	Cys	Ile	
				240					245				
GCC	CTC	CTT	GTG	GTC	GGC	ATC	ATG	TGT	GTG	GTG	GCC	TAC	1278
Ala	Leu	Leu	Val	Val	Gly	Ile	Met	Cys	Val	Val	Ala	Tyr	
	250					255					260		

FIG. 5B

20 / 29

TGC AAA ACC AAG AAA CAG CGG AAA AAG CTG CAT GAC CGT 1317
 Cys Lys Thr Lys Lys Gln Arg Lys Lys Leu His Asp Arg
 265 270

CTT CGG CAG AGC CTT CGG TCT GAA CGA AAC AAT ATG ATG 1356
 Leu Arg Gln Ser Leu Arg Ser Glu Arg Asn Asn Met Met
 275 280 285

AAC ATT GCC AAT GGG CCT CAC CAT CCT AAC CCA CCC CCC 1395
 Asn Ile Ala Asn Gly Pro His His Pro Asn Pro Pro Pro
 290 295 300

GAG AAT GTC CAG CTG GTG AAT CAA TAC GTA TCT AAA AAC 1434
 Glu Asn Val Gln Leu Val Asn Gln Tyr Val Ser Lys Asn
 305 310

GTC ATC TCC AGT GAG CAT ATT GTT GAG AGA GAA GCA GAG 1473
 Val Ile Ser Ser Glu His Ile Val Glu Arg Glu Ala Glu
 315 320 325

ACA TCC TTT TCC ACC AGT CAC TAT ACT TCC ACA GCC CAT 1512
 Thr Ser Phe Ser Thr Ser His Tyr Thr Ser Thr Ala His
 330 335

CAC TCC ACT ACT GTC ACC CAG ACT CCT AGC CAC AGC TGG 1551
 His Ser Thr Thr Val Thr Gln Thr Pro Ser His Ser Trp
 340 345 350

AGC AAC GGA CAC ACT GAA AGC ATC CTT TCC GAA AGC CAC 1590
 Ser Asn Gly His Thr Glu Ser Ile Leu Ser Glu Ser His
 355 360 365

TCT GTA ATC GTG ATG TCA TCC GTA GAA AAC AGT AGG CAC 1629
 Ser Val Ile Val Met Ser Ser Val Glu Asn Ser Arg His
 370 375

AGC AGC CCA ACT GGG GGC CCA AGA GGA CGT CTT AAT GGC 1668
 Ser Ser Pro Thr Gly Gly Pro Arg Gly Arg Leu Asn Gly
 380 385 390

ACA GGA GGC CCT CGT GAA TGT AAC AGC TTC CTC AGG CAT 1707
 Thr Gly Gly Pro Arg Glu Cys Asn Ser Phe Leu Arg His
 395 400

GCC AGA GAA ACC CCT GAT TCC TAC CGA GAC TCT CCT CAT 1746
 Ala Arg Glu Thr Pro Asp Ser Tyr Arg Asp Ser Pro His
 405 410 415

AGT GAA AGG TAAAA CCGAAGGCAA AGCTACTGCA GAGGAGAAAC 1790
 Ser Glu Arg
 420

FIG. 5C

21 / 29

TCAGTCAGAG AATCCCTGTG AGCACCTGCG GTCTCACCTC AGGAAATCTA 1840
CTCTAATCAG AATAAGGGGC GGCAGTTACC TGTTC TAGGA GTGCTCCTAG 1890
TTGATGAAGT CATCTCTTTG TTTGACGGAA CTTATTTCTT CTGAGCTTCT 1940
CTCGTCGTCC CAGTGACTGA CAGGCAACAG ACTCTTAAAG AGCTGGGATG 1990
CTTTGATGCG GAAGGTGCAG CACATGGAGT TTCCAGCTCT GGCCATGGGC 2040
TCAGACCCAC TCGGGGTCTC AGTGTCTCA GTTGTAACAT TAGAGAGATG 2090
GCATCAATGC TTGATAAGGA CCCTTCTATA ATTCCAATTG CCAGTTATCC 2140
AAACTCTGAT TCGGTGGTCG AGCTGGCCTC GTGTTCTTAT CTGCTAACCC 2190
TGTCTTACCT TCCAGCCTCA GTTAAGTCAA ATCAAGGGCT ATGTCATTGC 2240
TGAATGTCAT GGGGGGCAAC TGCTTGCCCT CCACCCTATA GTATCTATTT 2290
TATGAAATTC CAAGAAGGGA TGAATAAATA AATCTCTTGG ATGCTGCGTC 2340
TGGCAGTCTT CACGGGTGGT TTCAAAGCA GAAAAAAAAA AAAAAAAAAA 2390
AAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA A 2431

FIG._5D

16	1	M	S	E	R	K	E	G	R	G	K	G	K	K	E	R	G	S	G	K	K	P	E	S	A	A	G	S	Q	S	P	A	L	P	P	R	L	K	E	M	K	S	Q	E	S	A	A	G
11	1	M	S	E	R	K	E	G	R	G	K	G	K	K	E	R	G	S	G	K	K	P	E	S	A	A	G	S	Q	S	P	A	L	P	P	Q	L	K	E	M	K	S	Q	E	S	A	A	G
76	1	M	S	E	R	K	E	G	R	G	K	G	K	K	E	R	G	S	G	K	K	P	E	S	A	A	G	S	Q	S	P	A	L	P	P	Q	L	K	E	M	K	S	Q	E	S	A	A	G
84	1	M	S	E	R	K	E	G	R	G	K	G	K	K	E	R	G	S	G	K	K	P	E	S	A	A	G	S	Q	S	P	A	L	P	P	Q	L	K	E	M	K	S	Q	E	S	A	A	G
78	1	M	S	E	R	K	E	G	R	G	K	G	K	K	E	R	G	S	G	K	K	P	E	S	A	A	G	S	Q	S	P	A	L	P	P	Q	L	K	E	M	K	S	Q	E	S	A	A	G

16	51	S	K	L	V	L	R	C	E	T	S	S	E	Y	S	S	L	R	F	K	W	F	K	N	G	N	E	L	N	R	K	N	K	P	Q	N	I	K	I	Q	K	K	P	G	K	S	E	L	R	I	N
11	51	S	K	L	V	L	R	C	E	T	S	S	E	Y	S	S	L	R	F	K	W	F	K	N	G	N	E	L	N	R	K	N	K	P	Q	N	I	K	I	Q	K	K	P	G	K	S	E	L	R	I	N
76	51	S	K	L	V	L	R	C	E	T	S	S	E	Y	S	S	L	R	F	K	W	F	K	N	G	N	E	L	N	R	K	N	K	P	Q	N	I	K	I	Q	K	K	P	G	K	S	E	L	R	I	N
84	51	S	K	L	V	L	R	C	E	T	S	S	E	Y	S	S	L	R	F	K	W	F	K	N	G	N	E	L	N	R	K	N	K	P	Q	N	I	K	I	Q	K	K	P	G	K	S	E	L	R	I	N
78	51	S	K	L	V	L	R	C	E	T	S	S	E	Y	S	S	L	R	F	K	W	F	K	N	G	N	E	L	N	R	K	N	K	P	Q	N	I	K	I	Q	K	K	P	G	K	S	E	L	R	I	N

16	101	K	A	S	L	A	D	S	G	E	Y	M	C	K	V	I	S	K	L	G	N	D	S	A	S	A	N	I	T	I	V	E	S	N	E	I	I	T	G	M	P	A	S	T	E	G	A	Y	V	S	S
11	101	K	A	S	L	A	D	S	G	E	Y	M	C	K	V	I	S	K	L	G	N	D	S	A	S	A	N	I	T	I	V	E	S	N	E	I	I	T	G	M	P	A	S	T	E	G	A	Y	V	S	S
76	101	K	A	S	L	A	D	S	G	E	Y	M	C	K	V	I	S	K	L	G	N	D	S	A	S	A	N	I	T	I	V	E	S	N	E	I	I	T	G	M	P	A	S	T	E	G	A	Y	V	S	S
84	101	K	A	S	L	A	D	S	G	E	Y	M	C	K	V	I	S	K	L	G	N	D	S	A	S	A	N	I	T	I	V	E	S	N	E	I	I	T	G	M	P	A	S	T	E	G	A	Y	V	S	S
78	101	K	A	S	L	A	D	S	G	E	Y	M	C	K	V	I	S	K	L	G	N	D	S	A	S	A	N	I	T	I	V	E	S	N	E	I	I	T	G	M	P	A	S	T	E	G	A	Y	V	S	S

16	151	E	S	P	I	R	I	S	V	S	T	E	G	A	N	T	S	S	S	T	S	T	T	G	T	S	H	L	V	K	C	A	E	K	E	K	T	F	C	V	N	G	G	E	C	F	M	V	K
11	151	E	S	P	I	R	I	S	V	S	T	E	G	A	N	T	S	S	S	T	S	T	T	G	T	S	H	L	V	K	C	A	E	K	E	K	T	F	C	V	N	G	G	E	C	F	M	V	K
76	151	E	S	P	I	R	I	S	V	S	T	E	G	A	N	T	S	S	S	T	S	T	T	G	T	S	H	L	V	K	C	A	E	K	E	K	T	F	C	V	N	G	G	E	C	F	M	V	K
84	151	E	S	P	I	R	I	S	V	S	T	E	G	A	N	T	S	S	S	T	S	T	T	G	T	S	H	L	V	K	C	A	E	K	E	K	T	F	C	V	N	G	G	E	C	F	M	V	K
78	151	E	S	P	I	R	I	S	V	S	T	E	G	A	N	T	S	S	S	T	S	T	T	G	T	S	H	L	V	K	C	A	E	K	E	K	T	F	C	V	N	G	G	E	C	F	M	V	K

FIG.-6A

16	398	GGPRECN	SFLRHARE	TPDSYR	DSPHSERY	VSA	MTTPAR	MSPVDF	HTPSSSP
11	401	GGPRECN	SFLRHARE	TPDSYR	DSPHSERY	VSA	MTTPAR	MSPVDF	HTPSSSP
76	393	GGPRECN	SFLRHARE	TPDSYR	DSPHSERY	VSA	MTTPAR	MSPVDF	HTPSSSP
84	393	GGPRECN	SFLRHARE	TPDSYR	DSPHSER				
16	446	KSP	PPSEMS	PPVSSMT	VSMP	SMAS	VPFM	EEEE	RPLLVT
11	451	KSP	PPSEMS	8PPVSSMT	VSMP	SMAS	VPFM	EEEE	RPLLVT
76	443	KSP	PPSEMS	PPVSSMT	VS	KPS	MAVS	PFM	EEEE
16	498	QF	SSFH	HNPA	HD	SN	SLP	AS	PLRIVE
11	501	QF	SSFH	HNPA	HD	SN	SLP	AS	PLRIVE
76	493	QF	SSFH	HNPA	HD	SN	SLP	AS	PLRIVE
16	548	RA	KRTK	PNQ	HI	AN	RLE	VDS	NTSS
11	551	RA	KRTK	PNQ	HI	AN	RLE	VDS	NTSS
76	543	RA	KRTK	PNQ	HI	AN	RLE	VDS	NTSS
16	598	A	A	SLE	A	T	P	A	F
11	601	A	A	SLE	A	T	P	A	F
76	593	A	A	SLE	A	T	P	A	F

FIG.-6C

25 / 29

1 GGGTACCATGGGTCTGGTGAGCGCGTTTCCCGCCTGAGCGCAACTAGCGGC
51 GGGTCGTGGGCACCTCCAGAAAAGATCCCGCACCATCCTCCAGGATCCAA
101 TGGCCTTGGAGAGAGGGCTGCAGGGCCCACGGACATTGCTGACTCTTCAG
151 AACGTGCTGACATGGAGCCAGGTAGACTGAAATTATCATGTGTCCAAATT
201 AAAATTGCATACTTCAAGGATTATTTGAAGGACTATTCTTAGACCCTTTT
251 AAGAAGATTTAAAGAAAAACCACTCGGCCCTGAGTGCGGCGAGGACCCTG
301 TTTGTGGATGTGGAGGAGCGCGGGCCGGAGGCCATGGACGTGAAGGAGAG
1 M D V K E R
351 GAAGCCTTACCGCTCGCTGACCCGGCGCCGCGACGCCGAGCGCCGCTACA
7 K P Y R S L T R R R D A E R R Y T
401 CCAGCTCGTCCGCGGACAGCGAGGAGGGCAAAGCCCCGCAGAAATCGTAC
24 S S S A D S E E G K A P Q K S Y
451 AGCTCCAGCGAGACCCTGAAGGCCTACGACCAGGACGCCCGCCTAGCCTA
40 S S S E T L K A Y D Q D A R L A Y
501 TGGCAGCCGCGTCAAGGACATTGTGCCGACAGGAGGCCGAGGAATTCTGCC
57 G S R V K D I V P Q E A E E F C R
551 GCACAGGTGCCAACTTCACCCTGCGGGAGCTGGGGCTGGAAGAAGTAACG
74 T G A N F T L R E L G L E E V T
601 CCCCCTCACGGGACCCTGTACCGGACAGACATTGGCCTCCCCCACTGCGG
90 P P H G T L Y R T D I G L P H C G
651 CTA CTCCATGGGGGCTGGCTCTGATGCCGACATGGAGGCTGACACGGTGC
107 Y S M G A G S D A D M E A D T V L
701 TGTCCCCTGAGCACCCCGTGCGTCTGTGGGGCCGGAGCACACGGTCAGGG
124 S P E H P V R L W G R S T R S G
751 CGCAGCTCCTGCCTGTCCAGCCGGGCCAATTCCAATCTCACACTCACCGA
140 R S S C L S S R A N S N L T L T D
801 CACCGAGCATGAAAACACTGAGACTGATCATCCGGGCGGCCTGCAGAACC
157 T E H E N T E T D H P G G L Q N H
851 ACGCGCGGCTCCGGACGCGCGCCGCGCGCTCTCGCACGCCCACACCCCC
174 A R L R T P P P P L S H A H T P
901 AACCAGCACCACGCGGCCTCCATTA ACTCCCTGAACCGGGGCAACTTCAC
190 N Q H H A A S I N S L N R G N F T
951 GCCGAGGAGCAACCCCAGCCCGGCCCCCACGGACCACTCGCTCTCCGGAG
207 P R S N P S P A P T D H S L S G E
1001 AGCCCCCTGCCGGCGGCGCCAGGAGCCTGCCACGCCCAGGAGAACTGG
224 P P A G G A Q E P A H A Q E N W
1051 CTGCTCAACAGCAACATCCCCCTGGAGACCAGAAACCTAGGCAAGCAGCC
240 L L N S N I P L E T R N L G K Q P

FIG. 7A

26 / 29

1101 ATTCCTAGGGACATTGCAGGACAACCTCATTGAGATGGACATTCTCGGCG
257 F L G T L Q D N L I E M D I L G A

1151 CCTCCCGCCATGATGGGGCTTACAGTGACGGGCACTTCCTCTTCAAGCCT
274 S R H D G A Y S D G H F L F K P

1201 GGAGGCACCTCCCCGCTCTTCTGCACCACATCACCAGGGTACCCACTGAC
290 G G T S P L F C T T S P G Y P L T

1251 GTCCAGCACAGTGTACTCTCCTCCGCCCCGACCCCTGCCCCGCAGCACCT
307 S S T V Y S P P P R P L P R S T F

1301 TCGCCCGGCCGGCCTTTAACCTCAAGAAGCCCTCCAAGTACTGTAAGTGG
324 A R P A F N L K K P S K Y C N W

1351 AAGTGCGCAGCCCTGAGCGCCATCGTCATCTCAGCCACTCTGGTCATCCT
340 K C A A L S A I V I S A T L V I L

1401 GCTGGCATACTTTGTGGCCATGCACCTGTTTGGCCTAAACTGGCACCTGC
357 L A Y F V A M H L F G L N W H L Q

1451 AGCCGATGGAGGGGCAGATGTATGAGATCACGGAGGACACAGCCAGCAGT
374 P M E G Q M Y E I T E D T A S S

1501 TGGCCTGTGCCAACCGACGTCTCCCTATAACCCCTCAGGGGGCACTGGCTT
390 W P V P T D V S L Y P S G G T G L

1551 AGAGACCCCTGACAGGAAAGGCAAAGGAACACAGAAGGAAAGCCCAGTA
407 E T P D R K G K G T T E G K P S S

1601 GTTTCCTTTCCAGAGGACAGTTTCATAGATTCTGGAGAAATTGATGTGGGA
424 F F P E D S F I D S G E I D V G

1651 AGGCGAGCTTCCCAGAAGATTCTCCTGGCACTTTCTGGAGATCTCAAGT
440 R R A S Q K I P P G T F W R S Q V

1701 GTTCATAGACCATCCTGTGCATCTGAAATTCAATGTGTCTCTGGGAAAGG
457 F I D H P V H L K F N V S L G K A

1751 CAGCCCTGGTTGGCATTATGCGCAGAAAAGGCCTCCCTCCTTCACATACA
474 A L V G I Y G R K G L P P S H T

1801 CAGTTTGACTTTGTGGAGCTGCTGGATGGCAGGAGGCTCCTAACCCAGGA
490 Q F D F V E L L D G R R L L T Q E

1851 GGCGCGGAGCCTAGAGGGGACCCCGCGCCAGTCTCGGGGAACTGTGCCCC
507 A R S L E G T P R Q S R G T V P P

1901 CCTCCAGCCATGAGACAGGCTTCATCCAGTATTTGGATTCAAGGAATCTGG
524 S S H E T G F I Q Y L D S G I W

1951 CACTTGGCTTTTTTACAATGACGGAAAGGAGTCAGAAGTGGTTTCCTTTCT
540 H L A F Y N D G K E S E V V S F L

FIG. 7B

27 / 29

2001 CACCACTGCCATTGCCTTGCCTCCCCGATTGAAAGAGATGAAAAGCCAGG
557 T T A I A L P P R L K E M K S Q E

2051 AATCGGCTGCAGGTTCCAAACTAGTCCTTCGGTGTGAAACCAGTTCTGAA
574 S A A G S K L V L R C E T S S E

2101 TACTCCTCTCTCAGATTCAAGTGGTTCAAGAATGGGAATGAATTGAATCG
590 Y S S L R F K W F K N G N E L N R

2151 AAAAAACAAACCACAAAATATCAAGATACAAAAAAGCCAGGGAAGTCAG
607 K N K P Q N I K I Q K K P G K S E

2201 AACTTCGCATTAACAAAGCATCACTGGCTGATTCTGGAGAGTATATGTGC
624 L R I N K A S L A D S G E Y M C

2251 AAAGTGATCAGCAAATTAGGAAATGACAGTGCCTCTGCCAATATCACCAT
640 K V I S K L G N D S A S A N I T I

2301 CGTGGAAATCAAACGAGATCATCACTGGTATGCCAGCCTCAACTGAAGGAG
657 V E S N E I I T G M P A S T E G A

2351 CATATGTGTCTTCAGAGTCTCCCATTAGAATATCAGTATCCACAGAAGGA
674 Y V S S E S P I R I S V S T E G

2401 GCAAATACTTCTTCATCTACATCTACATCCACCCTGGGACAAGCCATCT
690 A N T S S S T S T S T T G T S H L

2451 TGTAATAATGTGCGGAGAAGGAGAAAACCTTTCTGTGTGAATGGAGGGGAGT
707 V K (C) A E K E K T F (C) V N G G E (C)

2501 GCTTCATGGTGAAGACCTTTCAAACCCCTCGAGATACTTGTGCAAGTGC
724 F M V K D L S N P S R Y L (C) K (C)

2551 CCAAATGAGTTTACTGGTGATCGCTGCCAAACTACGTAATGGCCAGCTT
740 P N E F T G D R (C) Q N Y V M A S F

2601 CTACAGTACGTCCACTCCCTTTCTGTCTCTGCCTGAATAGGAGCATGCTC
757 Y S T S T P F L S L P E

2651 AGTTGGTGCTGCTTTCTTGTTGCTGCATCTCCCCTCAGATTCCACCTAGA

2701 GCTAGATGTGTCTTACCAGATCTAATATTGACTGCCTCTGCCTGTGCGCAT
2751 GAGAACATTAACAAAAGCAATTGTATTACTTCCTCTGTTCGCGACTAGTT
2801 GGCTCTGAGATACTAATAGGTGTGTGAGGCTCCGGATGTTTCTGGAATTG
2851 ATATTGAATGATGTGATACAAATTGATAGTCAATATCAAGCAGTGAAATA
2901 TGATAATAAAGGCATTTCAAAGTCTCACTTTTATTGATAAAATAAAAATC
2951 ATTCTACTGAACAGTCCATCTTCTTTATACAAATGACCACATCCTGAAAAG
3001 GGTGTTGCTAAGCTGTAACCGATATGCACTTGAAATGATGGTAAGTTAAT
3051 TTTGATTTCAGAATGTGTTATTTGTACAAATAAACATAATAAAAGGAAAA
3101 AAAAAAAAAA

FIG. 7C

1 GAATTCCGGGACAGCCTCTCTCGCCCGCGTGTCTGCCCGCCGCGCCACCGCGGCTGGTGCTCTCTTCGCTTT

76 TACTTCTCTCGCATGACAGTTGTTTTCTTCATCTCGACAGACACCAGCTTCAGATGCTCGAGGTGAGAAACAATGC

1151 CTTTCAGTTTGGGCTACTGGTTTACTTAATTAAATCAGCCGGCAGCTCCGTGATCTATTTTCGTCCCCTGTCTCTCT

226 TGACGAGCCCGGATGGTTTGGAGTAGCAATTTAAAGAATAAAGTGGCCCCAGAAACAGCAGCTTTAAAGAAT

301 TATTACGATATACTTTTGATTTTGTAGTTGCTAGGAGCTTTTCTTCCCCCTTGCACTCTTCTGAACCTCTTCTTGA

376 TTTTAAATAGGCCCTTGGACTTGGACGATTTATCGATTTTCCCCCTGTAAAGATGCTGTATCATTTTGGTTGGGGGG

451 CCTCTGCGTGAATGGACCGTGAGAGCGGCCAGGCCCTTCTTCTGGAGGTGAGCCGATGGAGATTTATTTCCCCCAG
H E I Y S P D

526 ACATGTCTGAGTCCGCCGAGAGGTCTCTCAGCCCCCTCCACTCAGCTGAGTGCAGACCCCATCTCTTGTATGGGC
B H S E V A A E R S S S P S T Q L S A D P S L D G L

601 TTCCGGCAGCAGAAGACATGCCAGAGCCCCAGACTGAAGATGGAGAACCCCTGGACTGCTGGGCTGGCCGCTGC
33 P A A E D H P E P Q T E D G R T P G L V G L A V P

676 CCTGCTGTGCTAGAGCTGAGCGCCTGAGAGGTTGCTCAACTCAGAGNAAATCTGCATTTGTCCTCCCATCC
58 [C] [C] A [C] L E A E R L R G [C] L N S E K I [C] I V P I L

751 TGGCTTGCCCTGCTAGCCTCTGCCCTCTGCATCGCCCGCCTCAAGTGGGTATTTGTGACAAGATCTTTGAATATG
83 A [C] L V S L [C] L [C] I A G L K W V F V D K I F E Y D

826 ACTCTCCTACTCACCTTGACCCCTGGGGGTTAGCCAGGACCCTATATTTCTCTGGACGCAACTGCTGCCTCAG
108 S P T H L D P G G L G Q D P I I S L D A T A A S A

901 CTGTGTGGGTGCTGTAGGCATACACTTCACCTGTCTCTAGGGCTCAATCTGAAAGTGAGGTTCAAGTTACAG
133 V W V S S E A Y T S P V S R A Q S E S E V Q V T V

29 / 29

976 TGCAAGGTGACAAGGCTGTGTCTCTCTTTGAACCATCAGCGGCACCGACACCGAAGAAATCGTATTTTTCCTTTT
158 Q G D K A V V S F E P S A A P T P K N R I F A F S
1051 CTTTCTTGCCG1CCACTGCGCCATCCTTCCCTTACCCACCCGGAACCTGAGGTGAGAACGCCCAAGTCAGCAA
183 F L P S T A P S F P S P T R N P E V R T P K S A T
1126 CTCAGCCACAAACACAGAACTAATCTCCAACTGCTCCTNAACCTTTCTACATCTACATCCACCTGGACAA
208 Q P Q T T E T N L Q T A P K L S T S T T G T S
1201 GCCATCTTGTAATAATGTGGGAGAGAGAAACTTTCTGTGTGAATGGAGGGGAGTGCTTCTCATGTGAAAGACC
233 H L V K [C] A E K E K T F [C] V N G G E [C] F H V K D L
1276 TTYCAAACCCCTCGAGATACCTGTGCAAGTGCCCAATGAGTTTACTGGTGATCGCTGCCAAACTAGGTAATGG
258 S N P S R Y L [C] K [C] P N E F T G D R [C] Q N Y V H A
1351 CCAGCTTCTACAGTACGTCACCTCCCTTCTGTCTGTGCTGAATAGGAGCATGCTCAGTTGGTGCTGCTTCTT
283 S F Y S T S T P F L S L P E O
1426 GTTGTGCAATCTCCCTCAGATTCCACCTAGAGCTAGATGTGTCTTACCAGATCTAATATTGACTGCCCTCTGCCT
1501 GTGGCATGAGAACAATTAAACAAAGCAATTGTATTACTTCTCTGTTCGGGACTAGTTGGCTCTGAGATACTAATA
1576 GGTGTGTGAGGCTCCGGATGTTTCTCGAATTGATATTGAATGATGTGTATACAAATTGATAGTCMAATATCAAGCAG
1651 TGAATAATGATAATAAGGCATTTCAAAGTCTCACTTTTATTGATAAAATAAAATCATTTCTACTGAACAGTCCA
1726 TCTTCTTTATACAAATGACCACATCCTGTAAAGGGTGTGCTAAGCTGTAAACCGATATGCACCTTGAAATGATGGTA
1801 AGTTAATTTTGATTTCAGAAATGTGTTATTGTGCACAAATAAACATAATAAAAGGNAAAANAAACCCGAATTC

EGF
-like

FIG.-8B

Sequence Listing

<110> Genentech, Inc.

<120> Use of Heregulin as a Growth Factor

<130> P1145PCT

5 <141> 1999-02-03

<150> US 09/020,598

<151> 1998-02-04

<160> 14

<210> 1

10 <211> 669

<212> PRT

<213> Homo sapiens

<400> 1

15	Ala	Arg	Ala	Pro	Gln	Arg	Gly	Arg	Ser	Leu	Ser	Pro	Ser	Arg	Asp
	1				5					10					15
	Lys	Leu	Phe	Pro	Asn	Pro	Ile	Arg	Ala	Leu	Gly	Pro	Asn	Ser	Pro
					20					25					30
	Ala	Pro	Arg	Ala	Val	Arg	Val	Glu	Arg	Ser	Val	Ser	Gly	Glu	Met
					35					40					45
20	Ser	Glu	Arg	Lys	Glu	Gly	Arg	Gly	Lys	Gly	Lys	Gly	Lys	Lys	Lys
					50					55					60
	Glu	Arg	Gly	Ser	Gly	Lys	Lys	Pro	Glu	Ser	Ala	Ala	Gly	Ser	Gln
					65					70					75
	Ser	Pro	Ala	Leu	Pro	Pro	Arg	Leu	Lys	Glu	Met	Lys	Ser	Gln	Glu
25					80					85					90
	Ser	Ala	Ala	Gly	Ser	Lys	Leu	Val	Leu	Arg	Cys	Glu	Thr	Ser	Ser
					95					100					105
	Glu	Tyr	Ser	Ser	Leu	Arg	Phe	Lys	Trp	Phe	Lys	Asn	Gly	Asn	Glu
					110					115					120
30	Leu	Asn	Arg	Lys	Asn	Lys	Pro	Gln	Asn	Ile	Lys	Ile	Gln	Lys	Lys
					125					130					135
	Pro	Gly	Lys	Ser	Glu	Leu	Arg	Ile	Asn	Lys	Ala	Ser	Leu	Ala	Asp
					140					145					150
	Ser	Gly	Glu	Tyr	Met	Cys	Lys	Val	Ile	Ser	Lys	Leu	Gly	Asn	Asp
35					155					160					165
	Ser	Ala	Ser	Ala	Asn	Ile	Thr	Ile	Val	Glu	Ser	Asn	Glu	Ile	Ile

					170					175					180
					Thr Gly Met Pro Ala Ser Thr Glu Gly Ala Tyr Val Ser Ser Glu										
					185					190					195
5					Ser Pro Ile Arg Ile Ser Val Ser Thr Glu Gly Ala Asn Thr Ser										
					200					205					210
					Ser Ser Thr Ser Thr Ser Thr Thr Gly Thr Ser His Leu Val Lys										
					215					220					225
					Cys Ala Glu Lys Glu Lys Thr Phe Cys Val Asn Gly Gly Glu Cys										
					230					235					240
10					Phe Met Val Lys Asp Leu Ser Asn Pro Ser Arg Tyr Leu Cys Lys										
					245					250					255
					Cys Gln Pro Gly Phe Thr Gly Ala Arg Cys Thr Glu Asn Val Pro										
					260					265					270
15					Met Lys Val Gln Asn Gln Glu Lys Ala Glu Glu Leu Tyr Gln Lys										
					275					280					285
					Arg Val Leu Thr Ile Thr Gly Ile Cys Ile Ala Leu Leu Val Val										
					290					295					300
					Gly Ile Met Cys Val Val Ala Tyr Cys Lys Thr Lys Lys Gln Arg										
					305					310					315
20					Lys Lys Leu His Asp Arg Leu Arg Gln Ser Leu Arg Ser Glu Arg										
					320					325					330
					Asn Asn Met Met Asn Ile Ala Asn Gly Pro His His Pro Asn Pro										
					335					340					345
25					Pro Pro Glu Asn Val Gln Leu Val Asn Gln Tyr Val Ser Lys Asn										
					350					355					360
					Val Ile Ser Ser Glu His Ile Val Glu Arg Glu Ala Glu Thr Ser										
					365					370					375
					Phe Ser Thr Ser His Tyr Thr Ser Thr Ala His His Ser Thr Thr										
					380					385					390
30					Val Thr Gln Thr Pro Ser His Ser Trp Ser Asn Gly His Thr Glu										
					395					400					405
					Ser Ile Leu Ser Glu Ser His Ser Val Ile Val Met Ser Ser Val										
					410					415					420
35					Glu Asn Ser Arg His Ser Ser Pro Thr Gly Gly Pro Arg Gly Arg										
					425					430					435
					Leu Asn Gly Thr Gly Gly Pro Arg Glu Cys Asn Ser Phe Leu Arg										
					440					445					450

	His	Ala	Arg	Glu	Thr	Pro	Asp	Ser	Tyr	Arg	Asp	Ser	Pro	His	Ser	
					455					460					465	
	Glu	Arg	Tyr	Val	Ser	Ala	Met	Thr	Thr	Pro	Ala	Arg	Met	Ser	Pro	
					470					475					480	
5	Val	Asp	Phe	His	Thr	Pro	Ser	Ser	Pro	Lys	Ser	Pro	Pro	Ser	Glu	
					485					490					495	
	Met	Ser	Pro	Pro	Val	Ser	Ser	Met	Thr	Val	Ser	Met	Pro	Ser	Met	
					500					505					510	
10	Ala	Val	Ser	Pro	Phe	Met	Glu	Glu	Glu	Arg	Pro	Leu	Leu	Leu	Val	
					515					520					525	
	Thr	Pro	Pro	Arg	Leu	Arg	Glu	Lys	Lys	Phe	Asp	His	His	Pro	Gln	
					530					535					540	
	Gln	Phe	Ser	Ser	Phe	His	His	Asn	Pro	Ala	His	Asp	Ser	Asn	Ser	
					545					550					555	
15	Leu	Pro	Ala	Ser	Pro	Leu	Arg	Ile	Val	Glu	Asp	Glu	Glu	Tyr	Glu	
					560					565					570	
	Thr	Thr	Gln	Glu	Tyr	Glu	Pro	Ala	Gln	Glu	Pro	Val	Lys	Lys	Leu	
					575					580					585	
20	Ala	Asn	Ser	Arg	Arg	Ala	Lys	Arg	Thr	Lys	Pro	Asn	Gly	His	Ile	
					590					595					600	
	Ala	Asn	Arg	Leu	Glu	Val	Asp	Ser	Asn	Thr	Ser	Ser	Gln	Ser	Ser	
					605					610					615	
	Asn	Ser	Glu	Ser	Glu	Thr	Glu	Asp	Glu	Arg	Val	Gly	Glu	Asp	Thr	
					620					625					630	
25	Pro	Phe	Leu	Gly	Ile	Gln	Asn	Pro	Leu	Ala	Ala	Ser	Leu	Glu	Ala	
					635					640					645	
	Thr	Pro	Ala	Phe	Arg	Leu	Ala	Asp	Ser	Arg	Thr	Asn	Pro	Ala	Gly	
					650					655					660	
30	Arg	Phe	Ser	Thr	Gln	Glu	Glu	Ile	Gln							
					665				669							

<210> 2

<211> 2226

<212> DNA

<213> Homo sapiens

35 <400> 2

gg gcg cga gcg cct cag cgc ggc cgc tcg ctc tcc ccc 38

Ala Arg Ala Pro Gln Arg Gly Arg Ser Leu Ser Pro

1

5

10

tcg agg gac aaa ctt ttc cca aac ccg atc cga gcc ctt 77

	Ser	Arg	Asp	Lys	Leu	Phe	Pro	Asn	Pro	Ile	Arg	Ala	Leu	
			15					20					25	
	gga	cca	aac	tcg	cct	gcg	ccg	aga	gcc	gtc	cgc	gta	gag	116
5	Gly	Pro	Asn	Ser	Pro	Ala	Pro	Arg	Ala	Val	Arg	Val	Glu	
					30					35				
	cgc	tcc	gtc	tcc	ggc	gag	atg	tcc	gag	cgc	aaa	gaa	ggc	155
	Arg	Ser	Val	Ser	Gly	Glu	Met	Ser	Glu	Arg	Lys	Glu	Gly	
		40					45					50		
	aga	ggc	aaa	ggg	aag	ggc	aag	aag	aag	gag	cga	ggc	tcc	194
10	Arg	Gly	Lys	Gly	Lys	Gly	Lys	Lys	Lys	Glu	Arg	Gly	Ser	
				55					60					
	ggc	aag	aag	ccg	gag	tcc	gcg	gcg	ggc	agc	cag	agc	cca	233
	Gly	Lys	Lys	Pro	Glu	Ser	Ala	Ala	Gly	Ser	Gln	Ser	Pro	
		65				70					75			
15	gcc	ttg	cct	ccc	cga	ttg	aaa	gag	atg	aaa	agc	cag	gaa	272
	Ala	Leu	Pro	Pro	Arg	Leu	Lys	Glu	Met	Lys	Ser	Gln	Glu	
			80					85				90		
	tcg	gct	gca	ggt	tcc	aaa	cta	gtc	ctt	cgg	tgt	gaa	acc	311
	Ser	Ala	Ala	Gly	Ser	Lys	Leu	Val	Leu	Arg	Cys	Glu	Thr	
20					95					100				
	agt	tct	gaa	tac	tcc	tct	ctc	aga	ttc	aag	tgg	ttc	aag	350
	Ser	Ser	Glu	Tyr	Ser	Ser	Leu	Arg	Phe	Lys	Trp	Phe	Lys	
		105					110					115		
	aat	ggg	aat	gaa	ttg	aat	cga	aaa	aac	aaa	cca	caa	aat	389
25	Asn	Gly	Asn	Glu	Leu	Asn	Arg	Lys	Asn	Lys	Pro	Gln	Asn	
				120					125					
	atc	aag	ata	caa	aaa	aag	cca	ggg	aag	tca	gaa	ctt	cgc	428
	Ile	Lys	Ile	Gln	Lys	Lys	Pro	Gly	Lys	Ser	Glu	Leu	Arg	
		130				135					140			
30	att	aac	aaa	gca	tca	ctg	gct	gat	tct	gga	gag	tat	atg	467
	Ile	Asn	Lys	Ala	Ser	Leu	Ala	Asp	Ser	Gly	Glu	Tyr	Met	
			145					150					155	
	tgc	aaa	gtg	atc	agc	aaa	tta	gga	aat	gac	agt	gcc	tct	506
	Cys	Lys	Val	Ile	Ser	Lys	Leu	Gly	Asn	Asp	Ser	Ala	Ser	
35					160					165				
	gcc	aat	atc	acc	atc	gtg	gaa	tca	aac	gag	atc	atc	act	545
	Ala	Asn	Ile	Thr	Ile	Val	Glu	Ser	Asn	Glu	Ile	Ile	Thr	
		170					175					180		
	ggt	atg	cca	gcc	tca	act	gaa	gga	gca	tat	gtg	tct	tca	584
40	Gly	Met	Pro	Ala	Ser	Thr	Glu	Gly	Ala	Tyr	Val	Ser	Ser	
				185					190					
	gag	tct	ccc	att	aga	ata	tca	gta	tcc	aca	gaa	gga	gca	623

	Glu Ser Pro Ile Arg Ile Ser Val Ser Thr Glu Gly Ala	
	195	200 205
	aat act tct tca tct aca tct aca tcc acc act ggg aca 662	
5	Asn Thr Ser Ser Ser Thr Ser Thr Ser Thr Thr Gly Thr	210 215 220
	agc cat ctt gta aaa tgt gcg gag aag gag aaa act ttc 701	
	Ser His Leu Val Lys Cys Ala Glu Lys Glu Lys Thr Phe	225 230
	tgt gtg aat gga ggg gag tgc ttc atg gtg aaa gac ctt 740	
10	Cys Val Asn Gly Gly Glu Cys Phe Met Val Lys Asp Leu	235 240 245
	tca aac ccc tcg aga tac ttg tgc aag tgc caa cct gga 779	
	Ser Asn Pro Ser Arg Tyr Leu Cys Lys Cys Gln Pro Gly	250 255
15	ttc act gga gca aga tgt act gag aat gtg ccc atg aaa 818	
	Phe Thr Gly Ala Arg Cys Thr Glu Asn Val Pro Met Lys	260 265 270
	gtc caa aac caa gaa aag gcg gag gag ctg tac cag aag 857	
20	Val Gln Asn Gln Glu Lys Ala Glu Glu Leu Tyr Gln Lys	275 280 285
	aga gtg ctg acc ata acc ggc atc tgc atc gcc ctc ctt 896	
	Arg Val Leu Thr Ile Thr Gly Ile Cys Ile Ala Leu Leu	290 295
	gtg gtc ggc atc atg tgt gtg gtg gcc tac tgc aaa acc 935	
25	Val Val Gly Ile Met Cys Val Val Ala Tyr Cys Lys Thr	300 305 310
	aag aaa cag cgg aaa aag ctg cat gac cgt ctt cgg cag 974	
	Lys Lys Gln Arg Lys Lys Leu His Asp Arg Leu Arg Gln	315 320
30	agc ctt cgg tct gaa cga aac aat atg atg aac att gcc 1013	
	Ser Leu Arg Ser Glu Arg Asn Asn Met Met Asn Ile Ala	325 330 335
	aat ggg cct cac cat cct aac cca ccc ccc gag aat gtc 1052	
35	Asn Gly Pro His His Pro Asn Pro Pro Pro Glu Asn Val	340 345 350
	cag ctg gtg aat caa tac gta tct aaa aac gtc atc tcc 1091	
	Gln Leu Val Asn Gln Tyr Val Ser Lys Asn Val Ile Ser	355 360
	agt gag cat att gtt gag aga gaa gca gag aca tcc ttt 1130	
40	Ser Glu His Ile Val Glu Arg Glu Ala Glu Thr Ser Phe	365 370 375
	tcc acc agt cac tat act tcc aca gcc cat cac tcc act 1169	

	Ser	Thr	Ser	His	Tyr	Thr	Ser	Thr	Ala	His	His	Ser	Thr	
				380					385					
	act	gtc	acc	cag	act	cct	agc	cac	agc	tgg	agc	aac	gga	1208
	Thr	Val	Thr	Gln	Thr	Pro	Ser	His	Ser	Trp	Ser	Asn	Gly	
5	390					395					400			
	cac	act	gaa	agc	atc	ctt	tcc	gaa	agc	cac	tct	gta	atc	1247
	His	Thr	Glu	Ser	Ile	Leu	Ser	Glu	Ser	His	Ser	Val	Ile	
			405					410					415	
	gtg	atg	tca	tcc	gta	gaa	aac	agt	agg	cac	agc	agc	cca	1286
10	Val	Met	Ser	Ser	Val	Glu	Asn	Ser	Arg	His	Ser	Ser	Pro	
					420					425				
	act	ggg	ggc	cca	aga	gga	cgt	ctt	aat	ggc	aca	gga	ggc	1325
	Thr	Gly	Gly	Pro	Arg	Gly	Arg	Leu	Asn	Gly	Thr	Gly	Gly	
		430					435					440		
15	cct	cgt	gaa	tgt	aac	agc	ttc	ctc	agg	cat	gcc	aga	gaa	1364
	Pro	Arg	Glu	Cys	Asn	Ser	Phe	Leu	Arg	His	Ala	Arg	Glu	
				445					450					
	acc	cct	gat	tcc	tac	cga	gac	tct	cct	cat	agt	gaa	agg	1403
	Thr	Pro	Asp	Ser	Tyr	Arg	Asp	Ser	Pro	His	Ser	Glu	Arg	
20	455					460					465			
	tat	gtg	tca	gcc	atg	acc	acc	cgc	gct	cgt	atg	tca	cct	1442
	Tyr	Val	Ser	Ala	Met	Thr	Thr	Pro	Ala	Arg	Met	Ser	Pro	
			470					475					480	
	gta	gat	ttc	cac	acg	cca	agc	tcc	ccc	aaa	tcg	ccc	cct	1481
25	Val	Asp	Phe	His	Thr	Pro	Ser	Ser	Pro	Lys	Ser	Pro	Pro	
					485					490				
	tcg	gaa	atg	tct	cca	ccc	gtg	tcc	agc	atg	acg	gtg	tcc	1520
	Ser	Glu	Met	Ser	Pro	Pro	Val	Ser	Ser	Met	Thr	Val	Ser	
		495					500					505		
30	atg	cct	tcc	atg	gcg	gtc	agc	ccc	ttc	atg	gaa	gaa	gag	1559
	Met	Pro	Ser	Met	Ala	Val	Ser	Pro	Phe	Met	Glu	Glu	Glu	
				510					515					
	aga	cct	cta	ctt	ctc	gtg	aca	cca	cca	agg	ctg	cgg	gag	1598
	Arg	Pro	Leu	Leu	Leu	Val	Thr	Pro	Pro	Arg	Leu	Arg	Glu	
35	520					525					530			
	aag	aag	ttt	gac	cat	cac	cct	cag	cag	ttc	agc	tcc	ttc	1637
	Lys	Lys	Phe	Asp	His	His	Pro	Gln	Gln	Phe	Ser	Ser	Phe	
			535					540					545	
	cac	cac	aac	ccc	gcg	cat	gac	agt	aac	agc	ctc	cct	gct	1676
40	His	His	Asn	Pro	Ala	His	Asp	Ser	Asn	Ser	Leu	Pro	Ala	
					550					555				
	agc	ccc	ttg	agg	ata	gtg	gag	gat	gag	gag	tat	gaa	acg	1715

Ser Pro Leu Arg Ile Val Glu Asp Glu Glu Tyr Glu Thr
 560 565 570
 acc caa gag tac gag cca gcc caa gag cct gtt aag aaa 1754
 Thr Gln Glu Tyr Glu Pro Ala Gln Glu Pro Val Lys Lys
 5 575 580
 ctc gcc aat agc cgg cgg gcc aaa aga acc aag ccc aat 1793
 Leu Ala Asn Ser Arg Arg Ala Lys Arg Thr Lys Pro Asn
 585 590 595
 ggc cac att gct aac aga ttg gaa gtg gac agc aac aca 1832
 10 Gly His Ile Ala Asn Arg Leu Glu Val Asp Ser Asn Thr
 600 605 610
 agc tcc cag agc agt aac tca gag agt gaa aca gaa gat 1871
 Ser Ser Gln Ser Ser Asn Ser Glu Ser Glu Thr Glu Asp
 615 620
 gaa aga gta ggt gaa gat acg cct ttc ctg ggc ata cag 1910
 15 Glu Arg Val Gly Glu Asp Thr Pro Phe Leu Gly Ile Gln
 625 630 635
 aac ccc ctg gca gcc agt ctt gag gca aca cct gcc ttc 1949
 20 Asn Pro Leu Ala Ala Ser Leu Glu Ala Thr Pro Ala Phe
 640 645
 cgc ctg gct gac agc agg act aac cca gca ggc cgc ttc 1988
 Arg Leu Ala Asp Ser Arg Thr Asn Pro Ala Gly Arg Phe
 650 655 660
 tcg aca cag gaa gaa atc cag g ccaggctgtc tagtgtaatt 2030
 25 Ser Thr Gln Glu Glu Ile Gln
 665 669
 gctaaccaag accctattgc tgtataaaac cttaaataaac acatagattc 2080
 acctgtaaaa ctttatttta tataataaag tattccacct taaattaaac 2130
 aattttatttt atttttagcag ttctgcaaatt agaaaacagg aaaaaaactt 2180
 30 ttataaatta aatatatgta tgtaaaaatg aaaaaaaaaa aaaaaa 2226
 <210> 3
 <211> 675
 <212> PRT
 <213> Homo sapiens
 35 <400> 3
 Asp Lys Leu Phe Pro Asn Pro Ile Arg Ala Leu Gly Pro Asn Ser
 1 5 10 15
 Pro Ala Pro Arg Ala Val Arg Val Glu Arg Ser Val Ser Gly Glu
 20 25 30
 40 Met Ser Glu Arg Lys Glu Gly Arg Gly Lys Gly Lys Lys Lys

					35					40					45				
					Lys	Glu	Arg	Gly	Ser	Gly	Lys	Lys	Pro	Glu	Ser	Ala	Ala	Gly	Ser
					50					55								60	
5					Gln	Ser	Pro	Ala	Leu	Pro	Pro	Gln	Leu	Lys	Glu	Met	Lys	Ser	Gln
					65					70								75	
					Glu	Ser	Ala	Ala	Gly	Ser	Lys	Leu	Val	Leu	Arg	Cys	Glu	Thr	Ser
					80					85								90	
					Ser	Glu	Tyr	Ser	Ser	Leu	Arg	Phe	Lys	Trp	Phe	Lys	Asn	Gly	Asn
					95					100								105	
10					Glu	Leu	Asn	Arg	Lys	Asn	Lys	Pro	Gln	Asn	Ile	Lys	Ile	Gln	Lys
					110					115								120	
					Lys	Pro	Gly	Lys	Ser	Glu	Leu	Arg	Ile	Asn	Lys	Ala	Ser	Leu	Ala
					125					130								135	
15					Asp	Ser	Gly	Glu	Tyr	Met	Cys	Lys	Val	Ile	Ser	Lys	Leu	Gly	Asn
					140					145								150	
					Asp	Ser	Ala	Ser	Ala	Asn	Ile	Thr	Ile	Val	Glu	Ser	Asn	Glu	Ile
					155					160								165	
					Ile	Thr	Gly	Met	Pro	Ala	Ser	Thr	Glu	Gly	Ala	Tyr	Val	Ser	Ser
					170					175								180	
20					Glu	Ser	Pro	Ile	Arg	Ile	Ser	Val	Ser	Thr	Glu	Gly	Ala	Asn	Thr
					185					190								195	
					Ser	Ser	Ser	Thr	Ser	Thr	Ser	Thr	Thr	Gly	Thr	Ser	His	Leu	Val
					200					205								210	
25					Lys	Cys	Ala	Glu	Lys	Glu	Lys	Thr	Phe	Cys	Val	Asn	Gly	Gly	Glu
					215					220								225	
					Cys	Phe	Met	Val	Lys	Asp	Leu	Ser	Asn	Pro	Ser	Arg	Tyr	Leu	Cys
					230					235								240	
					Lys	Cys	Pro	Asn	Glu	Phe	Thr	Gly	Asp	Arg	Cys	Gln	Asn	Tyr	Val
					245					250								255	
30					Met	Ala	Ser	Phe	Tyr	Lys	His	Leu	Gly	Ile	Glu	Phe	Met	Glu	Ala
					260					265								270	
					Glu	Glu	Leu	Tyr	Gln	Lys	Arg	Val	Leu	Thr	Ile	Thr	Gly	Ile	Cys
					275					280								285	
35					Ile	Ala	Leu	Leu	Val	Val	Gly	Ile	Met	Cys	Val	Val	Ala	Tyr	Cys
					290					295								300	
					Lys	Thr	Lys	Lys	Gln	Arg	Lys	Lys	Leu	His	Asp	Arg	Leu	Arg	Gln
					305					310								315	

	Ser	Leu	Arg	Ser	Glu	Arg	Asn	Asn	Met	Met	Asn	Ile	Ala	Asn	Gly	
					320					325					330	
	Pro	His	His	Pro	Asn	Pro	Pro	Pro	Glu	Asn	Val	Gln	Leu	Val	Asn	
					335					340					345	
5	Gln	Tyr	Val	Ser	Lys	Asn	Val	Ile	Ser	Ser	Glu	His	Ile	Val	Glu	
					350					355					360	
	Arg	Glu	Ala	Glu	Thr	Ser	Phe	Ser	Thr	Ser	His	Tyr	Thr	Ser	Thr	
					365					370					375	
10	Ala	His	His	Ser	Thr	Thr	Val	Thr	Gln	Thr	Pro	Ser	His	Ser	Trp	
					380					385					390	
	Ser	Asn	Gly	His	Thr	Glu	Ser	Ile	Leu	Ser	Glu	Ser	His	Ser	Val	
					395					400					405	
	Ile	Val	Met	Ser	Ser	Val	Glu	Asn	Ser	Arg	His	Ser	Ser	Pro	Thr	
					410					415					420	
15	Gly	Gly	Pro	Arg	Gly	Arg	Leu	Asn	Gly	Thr	Gly	Gly	Pro	Arg	Glu	
					425					430					435	
	Cys	Asn	Ser	Phe	Leu	Arg	His	Ala	Arg	Glu	Thr	Pro	Asp	Ser	Tyr	
					440					445					450	
20	Arg	Asp	Ser	Pro	His	Ser	Glu	Arg	Tyr	Val	Ser	Ala	Met	Thr	Thr	
					455					460					465	
	Pro	Ala	Arg	Met	Ser	Pro	Val	Asp	Phe	His	Thr	Pro	Ser	Ser	Pro	
					470					475					480	
	Lys	Ser	Pro	Pro	Ser	Glu	Met	Ser	Pro	Pro	Val	Ser	Ser	Met	Thr	
					485					490					495	
25	Val	Ser	Met	Pro	Ser	Met	Ala	Val	Ser	Pro	Phe	Met	Glu	Glu	Glu	
					500					505					510	
	Arg	Pro	Leu	Leu	Leu	Val	Thr	Pro	Pro	Arg	Leu	Arg	Glu	Lys	Lys	
					515					520					525	
30	Phe	Asp	His	His	Pro	Gln	Gln	Phe	Ser	Ser	Phe	His	His	Asn	Pro	
					530					535					540	
	Ala	His	Asp	Ser	Asn	Ser	Leu	Pro	Ala	Ser	Pro	Leu	Arg	Ile	Val	
					545					550					555	
	Glu	Asp	Glu	Glu	Tyr	Glu	Thr	Thr	Gln	Glu	Tyr	Glu	Pro	Ala	Gln	
					560					565					570	
35	Glu	Pro	Val	Lys	Lys	Leu	Ala	Asn	Ser	Arg	Arg	Ala	Lys	Arg	Thr	
					575					580					585	
	Lys	Pro	Asn	Gly	His	Ile	Ala	Asn	Arg	Leu	Glu	Val	Asp	Ser	Asn	
					590					595					600	

Thr Ser Ser Gln Ser Ser Asn Ser Glu Ser Glu Thr Glu Asp Glu
 605 610 615
 Arg Val Gly Glu Asp Thr Pro Phe Leu Gly Ile Gln Asn Pro Leu
 620 625 630
 5 Ala Ala Ser Leu Glu Ala Thr Pro Ala Phe Arg Leu Ala Asp Ser
 635 640 645
 Arg Thr Asn Pro Ala Gly Arg Phe Ser Thr Gln Glu Glu Ile Gln
 650 655 660
 10 Ala Arg Leu Ser Ser Val Ile Ala Asn Gln Asp Pro Ile Ala Val
 665 670 675
 <210> 4
 <211> 2199
 <212> DNA
 <213> Homo sapiens
 15 <400> 4
 gg gac aaa ctt ttc cca aac ccg atc cga gcc ctt gga 38
 Asp Lys Leu Phe Pro Asn Pro Ile Arg Ala Leu Gly
 1 5 10
 cca aac tcg cct gcg ccg aga gcc gtc cgc gta gag cgc 77
 20 Pro Asn Ser Pro Ala Pro Arg Ala Val Arg Val Glu Arg
 15 20 25
 tcc gtc tcc ggc gag atg tcc gag cgc aaa gaa ggc aga 116
 Ser Val Ser Gly Glu Met Ser Glu Arg Lys Glu Gly Arg
 30 35
 25 ggc aaa ggg aag ggc aag aag aag gag cga gcc tcc ggc 155
 Gly Lys Gly Lys Gly Lys Lys Lys Glu Arg Gly Ser Gly
 40 45 50
 aag aag ccg gag tcc gcg gcg ggc agc cag agc cca gcc 194
 Lys Lys Pro Glu Ser Ala Ala Gly Ser Gln Ser Pro Ala
 30 55 60
 ttg cct ccc caa ttg aaa gag atg aaa agc cag gaa tcg 233
 Leu Pro Pro Gln Leu Lys Glu Met Lys Ser Gln Glu Ser
 65 70 75
 gct gca ggt tcc aaa cta gtc ctt cgg tgt gaa acc agt 272
 35 Ala Ala Gly Ser Lys Leu Val Leu Arg Cys Glu Thr Ser
 80 85 90
 tct gaa tac tcc tct ctc aga ttc aag tgg ttc aag aat 311
 Ser Glu Tyr Ser Ser Leu Arg Phe Lys Trp Phe Lys Asn
 95 100
 40 ggg aat gaa ttg aat cga aaa aac aaa cca caa aat atc 350
 Gly Asn Glu Leu Asn Arg Lys Asn Lys Pro Gln Asn Ile
 105 110 115

aag ata caa aaa aag cca ggg aag tca gaa ctt cgc att 389
 Lys Ile Gln Lys Lys Pro Gly Lys Ser Glu Leu Arg Ile
 120 125

5 aac aaa gca tca ctg gct gat tct gga gag tat atg tgc 428
 Asn Lys Ala Ser Leu Ala Asp Ser Gly Glu Tyr Met Cys
 130 135 140

aaa gtg atc agc aaa tta gga aat gac agt gcc tct gcc 467
 Lys Val Ile Ser Lys Leu Gly Asn Asp Ser Ala Ser Ala
 145 150 155

10 aat atc acc atc gtg gaa tca aac gag atc atc act ggt 506
 Asn Ile Thr Ile Val Glu Ser Asn Glu Ile Ile Thr Gly
 160 165

atg cca gcc tca act gaa gga gca tat gtg tct tca gag 545
 Met Pro Ala Ser Thr Glu Gly Ala Tyr Val Ser Ser Glu
 15 170 175 180

tct ccc att aga ata tca gta tcc aca gaa gga gca aat 584
 Ser Pro Ile Arg Ile Ser Val Ser Thr Glu Gly Ala Asn
 185 190

act tct tca tct aca tct aca tcc acc act ggg aca agc 623
 20 Thr Ser Ser Ser Thr Ser Thr Ser Thr Thr Gly Thr Ser
 195 200 205

cat ctt gta aaa tgt gcg gag aag gag aaa act ttc tgt 662
 His Leu Val Lys Cys Ala Glu Lys Glu Lys Thr Phe Cys
 210 215 220

25 gtg aat gga ggg gag tgc ttc atg gtg aaa gac ctt tca 701
 Val Asn Gly Gly Glu Cys Phe Met Val Lys Asp Leu Ser
 225 230

aac ccc tcg aga tac ttg tgc aag tgc cca aat gag ttt 740
 30 Asn Pro Ser Arg Tyr Leu Cys Lys Cys Pro Asn Glu Phe
 235 240 245

act ggt gat cgc tgc caa aac tac gta atg gcc agc ttc 779
 Thr Gly Asp Arg Cys Gln Asn Tyr Val Met Ala Ser Phe
 250 255

tac aag cat ctt ggg att gaa ttt atg gag gcg gag gag 818
 35 Tyr Lys His Leu Gly Ile Glu Phe Met Glu Ala Glu Glu
 260 265 270

ctg tac cag aag aga gtg ctg acc ata acc ggc atc tgc 857
 Leu Tyr Gln Lys Arg Val Leu Thr Ile Thr Gly Ile Cys
 275 280 285

40 atc gcc ctc ctt gtg gtc ggc atc atg tgt gtg gtg gcc 896
 Ile Ala Leu Leu Val Val Gly Ile Met Cys Val Val Ala
 290 295

	tac tgc aaa acc aag aaa cag cgg aaa aag ctg cat gac	935
	Tyr Cys Lys Thr Lys Lys Gln Arg Lys Lys Leu His Asp	
	300 305 310	
5	cgt ctt cgg cag agc ctt cgg tct gaa cga aac aat atg	974
	Arg Leu Arg Gln Ser Leu Arg Ser Glu Arg Asn Asn Met	
	315 320	
	atg aac att gcc aat ggg cct cac cat cct aac cca ccc	1013
	Met Asn Ile Ala Asn Gly Pro His His Pro Asn Pro Pro	
	325 330 335	
10	ccc gag aat gtc cag ctg gtg aat caa tac gta tct aaa	1052
	Pro Glu Asn Val Gln Leu Val Asn Gln Tyr Val Ser Lys	
	340 345 350	
	aac gtc atc tcc agt gag cat att gtt gag aga gaa gca	1091
15	Asn Val Ile Ser Ser Glu His Ile Val Glu Arg Glu Ala	
	355 360	
	gag aca tcc ttt tcc acc agt cac tat act tcc aca gcc	1130
	Glu Thr Ser Phe Ser Thr Ser His Tyr Thr Ser Thr Ala	
	365 370 375	
20	cat cac tcc act act gtc acc cag act cct agc cac agc	1169
	His His Ser Thr Thr Val Thr Gln Thr Pro Ser His Ser	
	380 385	
	tgg agc aac gga cac act gaa agc atc ctt tcc gaa agc	1208
	Trp Ser Asn Gly His Thr Glu Ser Ile Leu Ser Glu Ser	
	390 395 400	
25	cac tct gta atc gtg atg tca tcc gta gaa aac agt agg	1247
	His Ser Val Ile Val Met Ser Ser Val Glu Asn Ser Arg	
	405 410 415	
	cac agc agc cca act ggg ggc cca aga gga cgt ctt aat	1286
30	His Ser Ser Pro Thr Gly Gly Pro Arg Gly Arg Leu Asn	
	420 425	
	ggc aca gga ggc cct cgt gaa tgt aac agc ttc ctc agg	1325
	Gly Thr Gly Gly Pro Arg Glu Cys Asn Ser Phe Leu Arg	
	430 435 440	
35	cat gcc aga gaa acc cct gat tcc tac cga gac tct cct	1364
	His Ala Arg Glu Thr Pro Asp Ser Tyr Arg Asp Ser Pro	
	445 450	
	cat agt gaa agg tat gtg tca gcc atg acc acc ccg gct	1403
	His Ser Glu Arg Tyr Val Ser Ala Met Thr Thr Pro Ala	
	455 460 465	
40	cgt atg tca cct gta gat ttc cac acg cca agc tcc ccc	1442
	Arg Met Ser Pro Val Asp Phe His Thr Pro Ser Ser Pro	
	470 475 480	

```

aaa tcg ccc cct tcg gaa atg tct cca ccc gtg tcc agc 1481
Lys Ser Pro Pro Ser Glu Met Ser Pro Pro Val Ser Ser
                                485                                490

5   atg acg gtg tcc atg cct tcc atg gcg gtc agc ccc ttc 1520
    Met Thr Val Ser Met Pro Ser Met Ala Val Ser Pro Phe
                                495                                500                                505

    atg gaa gaa gag aga cct cta ctt ctc gtg aca cca cca 1559
    Met Glu Glu Glu Arg Pro Leu Leu Leu Val Thr Pro Pro
                                510                                515

10  agg ctg cgg gag aag aag ttt gac cat cac cct cag cag 1598
    Arg Leu Arg Glu Lys Lys Phe Asp His His Pro Gln Gln
    520                                525                                530

    ttc agc tcc ttc cac cac aac ccc gcg cat gac agt aac 1637
    Phe Ser Ser Phe His His Asn Pro Ala His Asp Ser Asn
15  535                                540                                545

    agc ctc cct gct agc ccc ttg agg ata gtg gag gat gag 1676
    Ser Leu Pro Ala Ser Pro Leu Arg Ile Val Glu Asp Glu
                                550                                555

20  gag tat gaa acg acc caa gag tac gag cca gcc caa gag 1715
    Glu Tyr Glu Thr Thr Gln Glu Tyr Glu Pro Ala Gln Glu
    560                                565                                570

    cct gtt aag aaa ctc gcc aat agc cgg cgg gcc aaa aga 1754
    Pro Val Lys Lys Leu Ala Asn Ser Arg Arg Ala Lys Arg
                                575                                580

25  acc aag ccc aat ggc cac att gct aac aga ttg gaa gtg 1793
    Thr Lys Pro Asn Gly His Ile Ala Asn Arg Leu Glu Val
    585                                590                                595

    gac agc aac aca agc tcc cag agc agt aac tca gag agt 1832
    Asp Ser Asn Thr Ser Ser Gln Ser Ser Asn Ser Glu Ser
30  600                                605                                610

    gaa aca gaa gat gaa aga gta ggt gaa gat acg cct ttc 1871
    Glu Thr Glu Asp Glu Arg Val Gly Glu Asp Thr Pro Phe
                                615                                620

    ctg ggc ata cag aac ccc ctg gca gcc agt ctt gag gca 1910
    Leu Gly Ile Gln Asn Pro Leu Ala Ala Ser Leu Glu Ala
35  625                                630                                635

    aca cct gcc ttc cgc ctg gct gac agc agg act aac cca 1949
    Thr Pro Ala Phe Arg Leu Ala Asp Ser Arg Thr Asn Pro
                                640                                645

40  gca ggc cgc ttc tcg aca cag gaa gaa atc cag gcc agg 1988
    Ala Gly Arg Phe Ser Thr Gln Glu Glu Ile Gln Ala Arg
    650                                655                                660

```


ctg tct agt gta att gct aac caa gac cct att gct gta ta 2029
 Leu Ser Ser Val Ile Ala Asn Gln Asp Pro Ile Ala Val
 665 670 675

a aacctaata aacacataga ttcacctgta aaactttatt 2070

5 ttatataata aagtattcca ccttaaatta aacaatttat tttatttttag 2120

cagttctgca aatagaaaac agggaaaaaaa cttttataaaa ttaaatatat 2170

gtatgtaaaa atgaaaaaaa aaaaaaaaaa 2199

<210> 5

<211> 637

10 <212> PRT

<213> Homo sapiens

<400> 5

Met Ser Glu Arg Lys Glu Gly Arg Gly Lys Gly Lys Gly Lys Lys
 1 5 10 15

15 Lys Glu Arg Gly Ser Gly Lys Lys Pro Glu Ser Ala Ala Gly Ser
 20 25 30

Gln Ser Pro Ala Leu Pro Pro Gln Leu Lys Glu Met Lys Ser Gln
 35 40 45

20 Glu Ser Ala Ala Gly Ser Lys Leu Val Leu Arg Cys Glu Thr Ser
 50 55 60

Ser Glu Tyr Ser Ser Leu Arg Phe Lys Trp Phe Lys Asn Gly Asn
 65 70 75

Glu Leu Asn Arg Lys Asn Lys Pro Gln Asn Ile Lys Ile Gln Lys
 80 85 90

25 Lys Pro Gly Lys Ser Glu Leu Arg Ile Asn Lys Ala Ser Leu Ala
 95 100 105

Asp Ser Gly Glu Tyr Met Cys Lys Val Ile Ser Lys Leu Gly Asn
 110 115 120

30 Asp Ser Ala Ser Ala Asn Ile Thr Ile Val Glu Ser Asn Glu Ile
 125 130 135

Ile Thr Gly Met Pro Ala Ser Thr Glu Gly Ala Tyr Val Ser Ser
 140 145 150

Glu Ser Pro Ile Arg Ile Ser Val Ser Thr Glu Gly Ala Asn Thr
 155 160 165

35 Ser Ser Ser Thr Ser Thr Ser Thr Thr Gly Thr Ser His Leu Val
 170 175 180

Lys Cys Ala Glu Lys Glu Lys Thr Phe Cys Val Asn Gly Gly Glu
 185 190 195

	Cys Phe Met Val Lys Asp Leu Ser Asn Pro Ser Arg Tyr Leu Cys	200	205	210
	Lys Cys Pro Asn Glu Phe Thr Gly Asp Arg Cys Gln Asn Tyr Val	215	220	225
5	Met Ala Ser Phe Tyr Lys Ala Glu Glu Leu Tyr Gln Lys Arg Val	230	235	240
	Leu Thr Ile Thr Gly Ile Cys Ile Ala Leu Leu Val Val Gly Ile	245	250	255
10	Met Cys Val Val Ala Tyr Cys Lys Thr Lys Lys Gln Arg Lys Lys	260	265	270
	Leu His Asp Arg Leu Arg Gln Ser Leu Arg Ser Glu Arg Asn Asn	275	280	285
	Met Met Asn Ile Ala Asn Gly Pro His His Pro Asn Pro Pro Pro	290	295	300
15	Glu Asn Val Gln Leu Val Asn Gln Tyr Val Ser Lys Asn Val Ile	305	310	315
	Ser Ser Glu His Ile Val Glu Arg Glu Ala Glu Thr Ser Phe Ser	320	325	330
20	Thr Ser His Tyr Thr Ser Thr Ala His His Ser Thr Thr Val Thr	335	340	345
	Gln Thr Pro Ser His Ser Trp Ser Asn Gly His Thr Glu Ser Ile	350	355	360
	Leu Ser Glu Ser His Ser Val Ile Val Met Ser Ser Val Glu Asn	365	370	375
25	Ser Arg His Ser Ser Pro Thr Gly Gly Pro Arg Gly Arg Leu Asn	380	385	390
	Gly Thr Gly Gly Pro Arg Glu Cys Asn Ser Phe Leu Arg His Ala	395	400	405
30	Arg Glu Thr Pro Asp Ser Tyr Arg Asp Ser Pro His Ser Glu Arg	410	415	420
	Tyr Val Ser Ala Met Thr Thr Pro Ala Arg Met Ser Pro Val Asp	425	430	435
	Phe His Thr Pro Ser Ser Pro Lys Ser Pro Pro Ser Glu Met Ser	440	445	450
35	Pro Pro Val Ser Ser Met Thr Val Ser Lys Pro Ser Met Ala Val	455	460	465
	Ser Pro Phe Met Glu Glu Glu Arg Pro Leu Leu Leu Val Thr Pro	470	475	480

	Pro Arg Leu Arg Glu Lys Lys Phe Asp His His Pro Gln Gln Phe	
	485	490 495
	Ser Ser Phe His His Asn Pro Ala His Asp Ser Asn Ser Leu Pro	
	500	505 510
5	Ala Ser Pro Leu Arg Ile Val Glu Asp Glu Glu Tyr Glu Thr Thr	
	515	520 525
	Gln Glu Tyr Glu Pro Ala Gln Glu Pro Val Lys Lys Leu Ala Asn	
	530	535 540
10	Ser Arg Arg Ala Lys Arg Thr Lys Pro Asn Gly His Ile Ala Asn	
	545	550 555
	Arg Leu Glu Val Asp Ser Asn Thr Ser Ser Gln Ser Ser Asn Ser	
	560	565 570
	Glu Ser Glu Thr Glu Asp Glu Arg Val Gly Glu Asp Thr Pro Phe	
	575	580 585
15	Leu Gly Ile Gln Asn Pro Leu Ala Ala Ser Leu Glu Ala Thr Pro	
	590	595 600
	Ala Phe Arg Leu Ala Asp Ser Arg Thr Asn Pro Ala Gly Arg Phe	
	605	610 615
20	Ser Thr Gln Glu Glu Ile Gln Ala Arg Leu Ser Ser Val Ile Ala	
	620	625 630
	Asn Gln Asp Pro Ile Ala Val	
	635 637	

<210> 6
 <211> 2490
 25 <212> DNA
 <213> Homo sapiens

<400> 6
 gtggctgcgg ggcaattgaa aaagagccgg cgaggagttc cccgaaactt 50
 gttggaactc cgggctcgcg cggaggccag gagctgagcg gcggcggtcg 100
 30 ccggacgatg ggagcgtgag caggacggtg ataactctc cccgatcggg 150
 ttgcgagggc gccgggcaga ggccaggacg cgagccgcca gcggcgggac 200
 ccatcgacga cttcccgggg cgacaggagc agccccgaga gccagggcga 250
 gcgcccgttc caggtggccg gaccgcccgc cgcgtccgcg ccgcgctccc 300
 tgcaggcaac gggagacgcc cccgcgcagc gcgagcgcct cagcgcggcc 350
 35 gctcgtctc cccatcgagg gacaaacttt tcccaaacc gatccgagcc 400
 cttggaccaa actcgctgc gccgagagcc gtccgcgtag agcgtccgt 450

```

ctccggcgag  atg tcc gag cgc aaa gaa ggc aga ggc aaa 490
Met Ser Glu Arg Lys Glu Gly Arg Gly Lys
      1              5              10

5  ggg aag ggc aag aag aag gag cga ggc tcc ggc aag aag 529
   Gly Lys Gly Lys Lys Lys Glu Arg Gly Ser Gly Lys Lys
               15              20

   ccg gag tcc gcg gcg ggc agc cag agc cca gcc ttg cct 568
   Pro Glu Ser Ala Ala Gly Ser Gln Ser Pro Ala Leu Pro
       25              30              35

10  ccc caa ttg aaa gag atg aaa agc cag gaa tcg gct gca 607
   Pro Gln Leu Lys Glu Met Lys Ser Gln Glu Ser Ala Ala
               40              45

   ggt tcc aaa cta gtc ctt cgg tgt gaa acc agt tct gaa 646
   Gly Ser Lys Leu Val Leu Arg Cys Glu Thr Ser Ser Glu
15   50              55              60

   tac tcc tct ctc aga ttc aag tgg ttc aag aat ggg aat 685
   Tyr Ser Ser Leu Arg Phe Lys Trp Phe Lys Asn Gly Asn
               65              70              75

20  gaa ttg aat cga aaa aac aaa cca caa aat atc aag ata 724
   Glu Leu Asn Arg Lys Asn Lys Pro Gln Asn Ile Lys Ile
               80              85

   caa aaa aag cca ggg aag tca gaa ctt cgc att aac aaa 763
   Gln Lys Lys Pro Gly Lys Ser Glu Leu Arg Ile Asn Lys
       90              95              100

25  gca tca ctg gct gat tct gga gag tat atg tgc aaa gtg 802
   Ala Ser Leu Ala Asp Ser Gly Glu Tyr Met Cys Lys Val
               105              110

   atc agc aaa tta gga aat gac agt gcc tct gcc aat atc 841
   Ile Ser Lys Leu Gly Asn Asp Ser Ala Ser Ala Asn Ile
30  115              120              125

   acc atc gtg gaa tca aac gag atc atc act ggt atg cca 880
   Thr Ile Val Glu Ser Asn Glu Ile Ile Thr Gly Met Pro
               130              135              140

   gcc tca act gaa gga gca tat gtg tct tca gag tct ccc 919
35  Ala Ser Thr Glu Gly Ala Tyr Val Ser Ser Glu Ser Pro
               145              150

   att aga ata tca gta tcc aca gaa gga gca aat act tct 958
   Ile Arg Ile Ser Val Ser Thr Glu Gly Ala Asn Thr Ser
               155              160              165

40  tca tct aca tct aca tcc acc act ggg aca agc cat ctt 997
   Ser Ser Thr Ser Thr Ser Thr Thr Gly Thr Ser His Leu
               170              175

```

gta aaa tgt gcg gag aag gag aaa act ttc tgt gtg aat 1036
 Val Lys Cys Ala Glu Lys Glu Lys Thr Phe Cys Val Asn
 180 185 190

5 gga ggg gag tgc ttc atg gtg aaa gac ctt tca aac ccc 1075
 Gly Gly Glu Cys Phe Met Val Lys Asp Leu Ser Asn Pro
 195 200 205

tcg aga tac ttg tgc aag tgc cca aat gag ttt act ggt 1114
 Ser Arg Tyr Leu Cys Lys Cys Pro Asn Glu Phe Thr Gly
 210 215

10 gat cgc tgc caa aac tac gta atg gcc agc ttc tac aag 1153
 Asp Arg Cys Gln Asn Tyr Val Met Ala Ser Phe Tyr Lys
 220 225 230

gcg gag gag ctg tac cag aag aga gtg ctg acc ata acc 1192
 15 Ala Glu Glu Leu Tyr Gln Lys Arg Val Leu Thr Ile Thr
 235 240

ggc atc tgc atc gcc ctc ctt gtg gtc ggc atc atg tgt 1231
 Gly Ile Cys Ile Ala Leu Leu Val Val Gly Ile Met Cys
 245 250 255

20 gtg gtg gcc tac tgc aaa acc aag aaa cag cgg aaa aag 1270
 Val Val Ala Tyr Cys Lys Thr Lys Lys Gln Arg Lys Lys
 260 265 270

ctg cat gac cgt ctt cgg cag agc ctt cgg tct gaa cga 1309
 Leu His Asp Arg Leu Arg Gln Ser Leu Arg Ser Glu Arg
 275 280

25 aac aat atg atg aac att gcc aat ggg cct cac cat cct 1348
 Asn Asn Met Met Asn Ile Ala Asn Gly Pro His His Pro
 285 290 295

aac cca ccc ccc gag aat gtc cag ctg gtg aat caa tac 1387
 30 Asn Pro Pro Pro Glu Asn Val Gln Leu Val Asn Gln Tyr
 300 305

gta tct aaa aac gtc atc tcc agt gag cat att gtt gag 1426
 Val Ser Lys Asn Val Ile Ser Ser Glu His Ile Val Glu
 310 315 320

35 aga gaa gca gag aca tcc ttt tcc acc agt cac tat act 1465
 Arg Glu Ala Glu Thr Ser Phe Ser Thr Ser His Tyr Thr
 325 330 335

tcc aca gcc cat cac tcc act act gtc acc cag act cct 1504
 Ser Thr Ala His His Ser Thr Thr Val Thr Gln Thr Pro
 340 345

40 agc cac agc tgg agc aac gga cac act gaa agc atc ctt 1543
 Ser His Ser Trp Ser Asn Gly His Thr Glu Ser Ile Leu
 350 355 360

	tcc gaa agc cac tct gta atc gtg atg tca tcc gta gaa	1582
	Ser Glu Ser His Ser Val Ile Val Met Ser Ser Val Glu	
	365 370	
5	aac agt agg cac agc agc cca act ggg ggc cca aga gga	1621
	Asn Ser Arg His Ser Ser Pro Thr Gly Gly Pro Arg Gly	
	375 380 385	
	cgt ctt aat ggc aca gga ggc cct cgt gaa tgt aac agc	1660
	Arg Leu Asn Gly Thr Gly Gly Pro Arg Glu Cys Asn Ser	
	390 395 400	
10	ttc ctc agg cat gcc aga gaa acc cct gat tcc tac cga	1699
	Phe Leu Arg His Ala Arg Glu Thr Pro Asp Ser Tyr Arg	
	405 410	
	gac tct cct cat agt gaa agg tat gtg tca gcc atg acc	1738
15	Asp Ser Pro His Ser Glu Arg Tyr Val Ser Ala Met Thr	
	415 420 425	
	acc ccg gct cgt atg tca cct gta gat ttc cac acg cca	1777
	Thr Pro Ala Arg Met Ser Pro Val Asp Phe His Thr Pro	
	430 435	
20	agc tcc ccc aaa tcg ccc cct tcg gaa atg tct cca ccc	1816
	Ser Ser Pro Lys Ser Pro Pro Ser Glu Met Ser Pro Pro	
	440 445 450	
	gtg tcc agc atg acg gtg tcc aag cct tcc atg gcg gtc	1855
	Val Ser Ser Met Thr Val Ser Lys Pro Ser Met Ala Val	
	455 460 465	
25	agc ccc ttc atg gaa gaa gag aga cct cta ctt ctc gtg	1894
	Ser Pro Phe Met Glu Glu Glu Arg Pro Leu Leu Leu Val	
	470 475	
	aca cca cca agg ctg cgg gag aag aag ttt gac cat cac	1933
30	Thr Pro Pro Arg Leu Arg Glu Lys Lys Phe Asp His His	
	480 485 490	
	cct cag cag ttc agc tcc ttc cac cac aac ccc gcg cat	1972
	Pro Gln Gln Phe Ser Ser Phe His His Asn Pro Ala His	
	495 500	
35	gac agt aac agc ctc cct gct agc ccc ttg agg ata gtg	2011
	Asp Ser Asn Ser Leu Pro Ala Ser Pro Leu Arg Ile Val	
	505 510 515	
	gag gat gag gag tat gaa acg acc caa gag tac gag cca	2050
	Glu Asp Glu Glu Tyr Glu Thr Thr Gln Glu Tyr Glu Pro	
	520 525 530	
40	gcc caa gag cct gtt aag aaa ctc gcc aat agc cgg cgg	2089
	Ala Gln Glu Pro Val Lys Lys Leu Ala Asn Ser Arg Arg	
	535 540	

```

gcc aaa aga acc aag ccc aat ggc cac att gct aac aga 2128
Ala Lys Arg Thr Lys Pro Asn Gly His Ile Ala Asn Arg
      545              550              555

      ttg gaa gtg gac agc aac aca agc tcc cag agc agt aac 2167
5    Leu Glu Val Asp Ser Asn Thr Ser Ser Gln Ser Ser Asn
      560              565

      tca gag agt gaa aca gaa gat gaa aga gta ggt gaa gat 2206
      Ser Glu Ser Glu Thr Glu Asp Glu Arg Val Gly Glu Asp
      570              575              580

10   acg cct ttc ctg ggc ata cag aac ccc ctg gca gcc agt 2245
      Thr Pro Phe Leu Gly Ile Gln Asn Pro Leu Ala Ala Ser
      585              590              595

      ctt gag gca aca cct gcc ttc cgc ctg gct gac agc agg 2284
15   Leu Glu Ala Thr Pro Ala Phe Arg Leu Ala Asp Ser Arg
      600              605

      act aac cca gca ggc cgc ttc tcg aca cag gaa gaa atc 2323
      Thr Asn Pro Ala Gly Arg Phe Ser Thr Gln Glu Glu Ile
      610              615              620

      cag gcc agg ctg tct agt gta att gct aac caa gac cct 2362
20   Gln Ala Arg Leu Ser Ser Val Ile Ala Asn Gln Asp Pro
      625              630

      att gct gta taaaacct aataaacaca tagattcacc tgtaaaactt 2410
      Ile Ala Val
      635      637

25   tattttatat aataaagtat tccaccttaa attaaacaat ttattttatt 2460

      ttagcagttc tgcaaataaa aaaaaaaaaa 2490

<210> 7
<211> 241
<212> PRT
30   <213> Homo sapiens

<400> 7
      Met Ser Glu Arg Lys Glu Gly Arg Gly Lys Gly Lys Gly Lys Lys
          1              5              10              15

      Lys Glu Arg Gly Ser Gly Lys Lys Pro Glu Ser Ala Ala Gly Ser
35   20              25              30

      Gln Ser Pro Ala Leu Pro Pro Gln Leu Lys Glu Met Lys Ser Gln
          35              40              45

      Glu Ser Ala Ala Gly Ser Lys Leu Val Leu Arg Cys Glu Thr Ser
          50              55              60

40   Ser Glu Tyr Ser Ser Leu Arg Phe Lys Trp Phe Lys Asn Gly Asn
          65              70              75

```

Glu Leu Asn Arg Lys Asn Lys Pro Gln Asn Ile Lys Ile Gln Lys
 80 85 90
 Lys Pro Gly Lys Ser Glu Leu Arg Ile Asn Lys Ala Ser Leu Ala
 95 100 105
 5 Asp Ser Gly Glu Tyr Met Cys Lys Val Ile Ser Lys Leu Gly Asn
 110 115 120
 Asp Ser Ala Ser Ala Asn Ile Thr Ile Val Glu Ser Asn Glu Ile
 125 130 135
 10 Ile Thr Gly Met Pro Ala Ser Thr Glu Gly Ala Tyr Val Ser Ser
 140 145 150
 Glu Ser Pro Ile Arg Ile Ser Val Ser Thr Glu Gly Ala Asn Thr
 155 160 165
 Ser Ser Ser Thr Ser Thr Ser Thr Thr Gly Thr Ser His Leu Val
 170 175 180
 15 Lys Cys Ala Glu Lys Glu Lys Thr Phe Cys Val Asn Gly Gly Glu
 185 190 195
 Cys Phe Met Val Lys Asp Leu Ser Asn Pro Ser Arg Tyr Leu Cys
 200 205 210
 20 Lys Cys Pro Asn Glu Phe Thr Gly Asp Arg Cys Gln Asn Tyr Val
 215 220 225
 Met Ala Ser Phe Tyr Ser Thr Ser Thr Pro Phe Leu Ser Leu Pro
 230 235 240
 Glu
 241
 25 <210> 8
 <211> 1715
 <212> DNA
 <213> Homo sapiens
 <400> 8
 30 gcgctgcct ccaacctgcg ggcgggaggt ggggtggtgc ggggcaattg 50
 aaaaagagcc ggcgaggagt tccccgaaac ttgttggaac tccgggctcg 100
 cgcggaggcc aggagctgag cggcggcggc tgccggacga tgggagcgtg 150
 agcaggacgg tgataacctc tccccgatcg ggttgcgagg gcgccgggca 200
 gaggccagga cgcgagccgc cagcggcggg acccatcgac gacttccccg 250
 35 ggcgacagga gcagccccga gagccagggc ggcgccccgt tccaggtggc 300
 cggaccgccc gccgcgtccg cgccgcgctc cctgcaggca acgggagacg 350


```

ccccgcgcga gcgcgagcgc ctcagcgcgg ccgctcgctc tccccatcga 400
gggacaaaact tttcccaaac ccgatccgag cccttggacc aaactcgctt 450
gcgccgagag ccgtccgcgt agagcgctcc gtctccggcg ag  atg 495
                                                Met
5                                                1

tcc gag cgc aaa gaa ggc aga ggc aaa ggg aag ggc aag 534
Ser Glu Arg Lys Glu Gly Arg Gly Lys Gly Lys Gly Lys
                    5                      10

aag aag gag cga ggc tcc ggc aag aag ccg gag tcc gcg 573
10 Lys Lys Glu Arg Gly Ser Gly Lys Lys Pro Glu Ser Ala
    15                      20                      25

gcg ggc agc cag agc cca gcc ttg cct ccc caa ttg aaa 612
Ala Gly Ser Gln Ser Pro Ala Leu Pro Pro Gln Leu Lys
                30                      35                      40

15 gag atg aaa agc cag gaa tcg gct gca ggt tcc aaa cta 651
Glu Met Lys Ser Gln Glu Ser Ala Ala Gly Ser Lys Leu
                45                      50

gtc ctt cgg tgt gaa acc agt tct gaa tac tcc tct ctc 690
20 Val Leu Arg Cys Glu Thr Ser Ser Glu Tyr Ser Ser Leu
    55                      60                      65

aga ttc aag tgg ttc aag aat ggg aat gaa ttg aat cga 729
Arg Phe Lys Trp Phe Lys Asn Gly Asn Glu Leu Asn Arg
                70                      75

aaa aac aaa cca caa aat atc aag ata caa aaa aag cca 768
25 Lys Asn Lys Pro Gln Asn Ile Lys Ile Gln Lys Lys Pro
    80                      85                      90

ggg aag tca gaa ctt cgc att aac aaa gca tca ctg gct 807
Gly Lys Ser Glu Leu Arg Ile Asn Lys Ala Ser Leu Ala
                95                      100                      105

30 gat tct gga gag tat atg tgc aaa gtg atc agc aaa tta 846
Asp Ser Gly Glu Tyr Met Cys Lys Val Ile Ser Lys Leu
                110                      115

gga aat gac agt gcc tct gcc aat atc acc atc gtg gaa 885
35 Gly Asn Asp Ser Ala Ser Ala Asn Ile Thr Ile Val Glu
    120                      125                      130

tca aac gag atc atc act ggt atg cca gcc tca act gaa 924
Ser Asn Glu Ile Ile Thr Gly Met Pro Ala Ser Thr Glu
                135                      140

gga gca tat gtg tct tca gag tct ccc att aga ata tca 963
40 Gly Ala Tyr Val Ser Ser Glu Ser Pro Ile Arg Ile Ser
    145                      150                      155

```

gta tcc aca gaa gga gca aat act tct tca tct aca tct 1002
 Val Ser Thr Glu Gly Ala Asn Thr Ser Ser Ser Thr Ser
 160 165 170

5 aca tcc acc act ggg aca agc cat ctt gta aaa tgt gcg 1041
 Thr Ser Thr Thr Gly Thr Ser His Leu Val Lys Cys Ala
 175 180

gag aag gag aaa act ttc tgt gtg aat gga ggg gag tgc 1080
 Glu Lys Glu Lys Thr Phe Cys Val Asn Gly Gly Glu Cys
 185 190 195

10 ttc atg gtg aaa gac ctt tca aac ccc tcg aga tac ttg 1119
 Phe Met Val Lys Asp Leu Ser Asn Pro Ser Arg Tyr Leu
 200 205

tgc aag tgc cca aat gag ttt act ggt gat cgc tgc caa 1158
 Cys Lys Cys Pro Asn Glu Phe Thr Gly Asp Arg Cys Gln
 15 210 215 220

aac tac gta atg gcc agc ttc tac agt acg tcc act ccc 1197
 Asn Tyr Val Met Ala Ser Phe Tyr Ser Thr Ser Thr Pro
 225 230 235

20 ttt ctg tct ctg cct gaa tagga gcatgctcag ttggtgctgc 1240
 Phe Leu Ser Leu Pro Glu
 240 241

tttcttggtg ctgcattctcc cctcagattc cacctagagc tagatgtgtc 1290

ttaccagatc taatattgac tgcctctgcc tgcgcgatga gaacattaac 1340

aaaagcaatt gtattacttc ctctgttcgc gactagttgg ctctgagata 1390

25 ctaatagggtg tgtgaggtc cggatgtttc tggaattgat attgaatgat 1440

gtgatacaaa ttgatagtca atatcaagca gtgaaatatg ataataaagg 1490

catttcaaag tctcactttt attgataaaa taaaaatcat tctactgaac 1540

agtccatctt ctttatacaa tgaccacatc ctgaaaaggg tgttgctaa 1590

ctgtaaccga tatgcacttg aaatgatggt aagttaattt tgattcagaa 1640

30 tgtgttattt gtcacaaata aacataataa aaggagttca gatgtttttc 1690

ttcattaacc aaaaaaaaaa aaaaa 1715

<210> 9
 <211> 420
 <212> PRT
 35 <213> Homo sapiens

<400> 9
 Met Ser Glu Arg Lys Glu Gly Arg Gly Lys Gly Lys Gly Lys Lys
 1 5 10 15

	Lys	Glu	Arg	Gly	Ser	Gly	Lys	Lys	Pro	Glu	Ser	Ala	Ala	Gly	Ser	
					20					25					30	
	Gln	Ser	Pro	Ala	Leu	Pro	Pro	Gln	Leu	Lys	Glu	Met	Lys	Ser	Gln	
					35					40					45	
5	Glu	Ser	Ala	Ala	Gly	Ser	Lys	Leu	Val	Leu	Arg	Cys	Glu	Thr	Ser	
					50					55					60	
	Ser	Glu	Tyr	Ser	Ser	Leu	Arg	Phe	Lys	Trp	Phe	Lys	Asn	Gly	Asn	
					65					70					75	
10	Glu	Leu	Asn	Arg	Lys	Asn	Lys	Pro	Gln	Asn	Ile	Lys	Ile	Gln	Lys	
					80					85					90	
	Lys	Pro	Gly	Lys	Ser	Glu	Leu	Arg	Ile	Asn	Lys	Ala	Ser	Leu	Ala	
					95					100					105	
	Asp	Ser	Gly	Glu	Tyr	Met	Cys	Lys	Val	Ile	Ser	Lys	Leu	Gly	Asn	
					110					115					120	
15	Asp	Ser	Ala	Ser	Ala	Asn	Ile	Thr	Ile	Val	Glu	Ser	Asn	Glu	Ile	
					125					130					135	
	Ile	Thr	Gly	Met	Pro	Ala	Ser	Thr	Glu	Gly	Ala	Tyr	Val	Ser	Ser	
					140					145					150	
20	Glu	Ser	Pro	Ile	Arg	Ile	Ser	Val	Ser	Thr	Glu	Gly	Ala	Asn	Thr	
					155					160					165	
	Ser	Ser	Ser	Thr	Ser	Thr	Ser	Thr	Thr	Gly	Thr	Ser	His	Leu	Val	
					170					175					180	
	Lys	Cys	Ala	Glu	Lys	Glu	Lys	Thr	Phe	Cys	Val	Asn	Gly	Gly	Glu	
					185					190					195	
25	Cys	Phe	Met	Val	Lys	Asp	Leu	Ser	Asn	Pro	Ser	Arg	Tyr	Leu	Cys	
					200					205					210	
	Lys	Cys	Pro	Asn	Glu	Phe	Thr	Gly	Asp	Arg	Cys	Gln	Asn	Tyr	Val	
					215					220					225	
30	Met	Ala	Ser	Phe	Tyr	Lys	Ala	Glu	Glu	Leu	Tyr	Gln	Lys	Arg	Val	
					230					235					240	
	Leu	Thr	Ile	Thr	Gly	Ile	Cys	Ile	Ala	Leu	Leu	Val	Val	Gly	Ile	
					245					250					255	
	Met	Cys	Val	Val	Ala	Tyr	Cys	Lys	Thr	Lys	Lys	Gln	Arg	Lys	Lys	
					260					265					270	
35	Leu	His	Asp	Arg	Leu	Arg	Gln	Ser	Leu	Arg	Ser	Glu	Arg	Asn	Asn	
					275					280					285	
	Met	Met	Asn	Ile	Ala	Asn	Gly	Pro	His	His	Pro	Asn	Pro	Pro	Pro	
					290					295					300	

Glu Asn Val Gln Leu Val Asn Gln Tyr Val Ser Lys Asn Val Ile
 305 310 315
 Ser Ser Glu His Ile Val Glu Arg Glu Ala Glu Thr Ser Phe Ser
 320 325 330
 5 Thr Ser His Tyr Thr Ser Thr Ala His His Ser Thr Thr Val Thr
 335 340 345
 Gln Thr Pro Ser His Ser Trp Ser Asn Gly His Thr Glu Ser Ile
 350 355 360
 10 Leu Ser Glu Ser His Ser Val Ile Val Met Ser Ser Val Glu Asn
 365 370 375
 Ser Arg His Ser Ser Pro Thr Gly Gly Pro Arg Gly Arg Leu Asn
 380 385 390
 Gly Thr Gly Gly Pro Arg Glu Cys Asn Ser Phe Leu Arg His Ala
 395 400 405
 15 Arg Glu Thr Pro Asp Ser Tyr Arg Asp Ser Pro His Ser Glu Arg
 410 415 420
 <210> 10
 <211> 2431
 <212> DNA
 20 <213> Homo sapiens
 <400> 10
 gaggcgcctg cctccaacct gcgggcggga ggtgggtggc tgcggggcaa 50
 ttgaaaaaga gccggcgagg agttccccga aacttggttg aactccgggc 100
 tcgcgaggag gccaggagct gagcggcggc ggctgccgga cgatgggagc 150
 25 gtgagcagga cggtgataac ctctccccga tcgggttgcg agggcgccgg 200
 gcagaggcca ggacgcgagc cgccagcggc gggacccatc gacgacttcc 250
 cggggcgaca ggagcagccc cgagagccag ggcgagcgcc cgttccaggt 300
 ggccggaccg cccgccgcgt ccgcgccgcg ctccctgcag gcaacgggag 350
 acgccccgcg gcagcgcgag cgctcagcg cggccgctcg ctctcccat 400
 30 cgagggacaa acttttccca aaccgatcc gagcccttgg accaaactcg 450
 cctgcgccga gagccgtccg cgtagagcgc tccgtctccg gcgag atg 498
 Met
 1
 tcc gag cgc aaa gaa ggc aga ggc aaa ggg aag ggc aag 537
 35 Ser Glu Arg Lys Glu Gly Arg Gly Lys Gly Lys Gly Lys
 5 10

aag aag gag cga ggc tcc ggc aag aag ccg gag tcc gcg 576
 Lys Lys Glu Arg Gly Ser Gly Lys Lys Pro Glu Ser Ala
 15 20 25

gcg ggc agc cag agc cca gcc ttg cct ccc caa ttg aaa 615
 5 Ala Gly Ser Gln Ser Pro Ala Leu Pro Pro Gln Leu Lys
 30 35 40

gag atg aaa agc cag gaa tcg gct gca ggt tcc aaa cta 654
 Glu Met Lys Ser Gln Glu Ser Ala Ala Gly Ser Lys Leu
 45 50

gtc ctt cgg tgt gaa acc agt tct gaa tac tcc tct ctc 693
 10 Val Leu Arg Cys Glu Thr Ser Ser Glu Tyr Ser Ser Leu
 55 60 65

aga ttc aag tgg ttc aag aat ggg aat gaa ttg aat cga 732
 15 Arg Phe Lys Trp Phe Lys Asn Gly Asn Glu Leu Asn Arg
 70 75

aaa aac aaa cca caa aat atc aag ata caa aaa aag cca 771
 Lys Asn Lys Pro Gln Asn Ile Lys Ile Gln Lys Lys Pro
 80 85 90

ggg aag tca gaa ctt cgc att aac aaa gca tca ctg gct 810
 20 Gly Lys Ser Glu Leu Arg Ile Asn Lys Ala Ser Leu Ala
 95 100 105

gat tct gga gag tat atg tgc aaa gtg atc agc aaa tta 849
 Asp Ser Gly Glu Tyr Met Cys Lys Val Ile Ser Lys Leu
 110 115

gga aat gac agt gcc tct gcc aat atc acc atc gtg gaa 888
 25 Gly Asn Asp Ser Ala Ser Ala Asn Ile Thr Ile Val Glu
 120 125 130

tca aac gag atc atc act ggt atg cca gcc tca act gaa 927
 30 Ser Asn Glu Ile Ile Thr Gly Met Pro Ala Ser Thr Glu
 135 140

gga gca tat gtg tct tca gag tct ccc att aga ata tca 966
 Gly Ala Tyr Val Ser Ser Glu Ser Pro Ile Arg Ile Ser
 145 150 155

gta tcc aca gaa gga gca aat act tct tca tct aca tct 1005
 35 Val Ser Thr Glu Gly Ala Asn Thr Ser Ser Ser Thr Ser
 160 165 170

aca tcc acc act ggg aca agc cat ctt gta aaa tgt gcg 1044
 Thr Ser Thr Thr Gly Thr Ser His Leu Val Lys Cys Ala
 175 180

gag aag gag aaa act ttc tgt gtg aat gga ggg gag tgc 1083
 40 Glu Lys Glu Lys Thr Phe Cys Val Asn Gly Gly Glu Cys
 185 190 195

	ttc atg gtg aaa gac ctt tca aac ccc tcg aga tac ttg 1122
	Phe Met Val Lys Asp Leu Ser Asn Pro Ser Arg Tyr Leu
	200 205
5	tgc aag tgt cca aat gag ttt act ggt gat cgc tgc caa 1161
	Cys Lys Cys Pro Asn Glu Phe Thr Gly Asp Arg Cys Gln
	210 215 220
	aac tac gta atg gcc agc ttc tac aag gcg gag gag ctg 1200
	Asn Tyr Val Met Ala Ser Phe Tyr Lys Ala Glu Glu Leu
	225 230 235
10	tac cag aag aga gtg ctg acc ata acc ggc atc tgc atc 1239
	Tyr Gln Lys Arg Val Leu Thr Ile Thr Gly Ile Cys Ile
	240 245
	gcc ctc ctt gtg gtc ggc atc atg tgt gtg gtg gcc tac 1278
	Ala Leu Leu Val Val Gly Ile Met Cys Val Val Ala Tyr
15	250 255 260
	tgc aaa acc aag aaa cag cgg aaa aag ctg cat gac cgt 1317
	Cys Lys Thr Lys Lys Gln Arg Lys Lys Leu His Asp Arg
	265 270
	ctt cgg cag agc ctt cgg tct gaa cga aac aat atg atg 1356
20	Leu Arg Gln Ser Leu Arg Ser Glu Arg Asn Asn Met Met
	275 280 285
	aac att gcc aat ggg cct cac cat cct aac cca ccc ccc 1395
	Asn Ile Ala Asn Gly Pro His His Pro Asn Pro Pro Pro
	290 295 300
25	gag aat gtc cag ctg gtg aat caa tac gta tct aaa aac 1434
	Glu Asn Val Gln Leu Val Asn Gln Tyr Val Ser Lys Asn
	305 310
	gtc atc tcc agt gag cat att gtt gag aga gaa gca gag 1473
	Val Ile Ser Ser Glu His Ile Val Glu Arg Glu Ala Glu
30	315 320 325
	aca tcc ttt tcc acc agt cac tat act tcc aca gcc cat 1512
	Thr Ser Phe Ser Thr Ser His Tyr Thr Ser Thr Ala His
	330 335
	cac tcc act act gtc acc cag act cct agc cac agc tgg 1551
35	His Ser Thr Thr Val Thr Gln Thr Pro Ser His Ser Trp
	340 345 350
	agc aac gga cac act gaa agc atc ctt tcc gaa agc cac 1590
	Ser Asn Gly His Thr Glu Ser Ile Leu Ser Glu Ser His
	355 360 365
40	tct gta atc gtg atg tca tcc gta gaa aac agt agg cac 1629
	Ser Val Ile Val Met Ser Ser Val Glu Asn Ser Arg His
	370 375

agc agc cca act ggg ggc cca aga gga cgt ctt aat ggc 1668
 Ser Ser Pro Thr Gly Gly Pro Arg Gly Arg Leu Asn Gly
 380 385 390

aca gga ggc cct cgt gaa tgt aac agc ttc ctc agg cat 1707
 5 Thr Gly Gly Pro Arg Glu Cys Asn Ser Phe Leu Arg His
 395 400

gcc aga gaa acc cct gat tcc tac cga gac tct cct cat 1746
 Ala Arg Glu Thr Pro Asp Ser Tyr Arg Asp Ser Pro His
 405 410 415

10 agt gaa agg taaaa ccgaaggcaa agctactgca gaggagaaac 1790
 Ser Glu Arg
 420

tcagtcagag aatccctgtg agcacctgcg gtctcacctc aggaaatcta 1840
 ctctaatacag aataaggggc ggcagttacc tgttcttagga gtgctcctag 1890
 15 ttgatgaagt catctctttg tttgacggaa cttatttctt ctgagcttct 1940
 ctcgctcgtcc cagtgactga caggcaacag actcttaaag agctgggatg 1990
 ctttgatgcg gaaggtgcag cacatggagt ttccagctct ggccatgggc 2040
 tcagaccac tcggggtctc agtgtcctca gttgtaacat tagagagatg 2090
 gcatcaatgc ttgataagga cccttctata attccaattg ccagttatcc 2140
 20 aaactctgat tcggtggctg agctggcctc gtgttcttat ctgctaacc 2190
 tgtcttacct tccagcctca gttaagtcaa atcaagggt atgtcattgc 2240
 tgaatgtcat ggggggcaac tgcttgccct ccacctata gtatctatct 2290
 tatgaaattc caagaaggga tgaataaata aatctcttgg atgctgcgtc 2340
 tggcagtctt cacgggtggt tttcaaagca gaaaaaaaaa aaaaaaaaaa 2390
 25 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa a 2431

<210> 11
 <211> 768
 <212> PRT
 <213> Homo sapiens

30 <400> 11
 Met Asp Val Lys Glu Arg Lys Pro Tyr Arg Ser Leu Thr Arg Arg
 1 5 10 15
 Arg Asp Ala Glu Arg Arg Tyr Thr Ser Ser Ser Ala Asp Ser Glu
 20 25 30

35 Glu Gly Lys Ala Pro Gln Lys Ser Tyr Ser Ser Ser Glu Thr Leu
 35 40 45

	Lys	Ala	Tyr	Asp	Gln	Asp	Ala	Arg	Leu	Ala	Tyr	Gly	Ser	Arg	Val	50	55	60
	Lys	Asp	Ile	Val	Pro	Gln	Glu	Ala	Glu	Glu	Phe	Cys	Arg	Thr	Gly	65	70	75
5	Ala	Asn	Phe	Thr	Leu	Arg	Glu	Leu	Gly	Leu	Glu	Glu	Val	Thr	Pro	80	85	90
	Pro	His	Gly	Thr	Leu	Tyr	Arg	Thr	Asp	Ile	Gly	Leu	Pro	His	Cys	95	100	105
10	Gly	Tyr	Ser	Met	Gly	Ala	Gly	Ser	Asp	Ala	Asp	Met	Glu	Ala	Asp	110	115	120
	Thr	Val	Leu	Ser	Pro	Glu	His	Pro	Val	Arg	Leu	Trp	Gly	Arg	Ser	125	130	135
	Thr	Arg	Ser	Gly	Arg	Ser	Ser	Cys	Leu	Ser	Ser	Arg	Ala	Asn	Ser	140	145	150
15	Asn	Leu	Thr	Leu	Thr	Asp	Thr	Glu	His	Glu	Asn	Thr	Glu	Thr	Asp	155	160	165
	His	Pro	Gly	Gly	Leu	Gln	Asn	His	Ala	Arg	Leu	Arg	Thr	Pro	Pro	170	175	180
20	Pro	Pro	Leu	Ser	His	Ala	His	Thr	Pro	Asn	Gln	His	His	Ala	Ala	185	190	195
	Ser	Ile	Asn	Ser	Leu	Asn	Arg	Gly	Asn	Phe	Thr	Pro	Arg	Ser	Asn	200	205	210
	Pro	Ser	Pro	Ala	Pro	Thr	Asp	His	Ser	Leu	Ser	Gly	Glu	Pro	Pro	215	220	225
25	Ala	Gly	Gly	Ala	Gln	Glu	Pro	Ala	His	Ala	Gln	Glu	Asn	Trp	Leu	230	235	240
	Leu	Asn	Ser	Asn	Ile	Pro	Leu	Glu	Thr	Arg	Asn	Leu	Gly	Lys	Gln	245	250	255
30	Pro	Phe	Leu	Gly	Thr	Leu	Gln	Asp	Asn	Leu	Ile	Glu	Met	Asp	Ile	260	265	270
	Leu	Gly	Ala	Ser	Arg	His	Asp	Gly	Ala	Tyr	Ser	Asp	Gly	His	Phe	275	280	285
	Leu	Phe	Lys	Pro	Gly	Gly	Thr	Ser	Pro	Leu	Phe	Cys	Thr	Thr	Ser	290	295	300
35	Pro	Gly	Tyr	Pro	Leu	Thr	Ser	Ser	Thr	Val	Tyr	Ser	Pro	Pro	Pro	305	310	315
	Arg	Pro	Leu	Pro	Arg	Ser	Thr	Phe	Ala	Arg	Pro	Ala	Phe	Asn	Leu	320	325	330

	Lys	Lys	Pro	Ser	Lys	Tyr	Cys	Asn	Trp	Lys	Cys	Ala	Ala	Leu	Ser	
					335					340					345	
	Ala	Ile	Val	Ile	Ser	Ala	Thr	Leu	Val	Ile	Leu	Leu	Ala	Tyr	Phe	
					350					355					360	
5	Val	Ala	Met	His	Leu	Phe	Gly	Leu	Asn	Trp	His	Leu	Gln	Pro	Met	
					365					370					375	
	Glu	Gly	Gln	Met	Tyr	Glu	Ile	Thr	Glu	Asp	Thr	Ala	Ser	Ser	Trp	
					380					385					390	
10	Pro	Val	Pro	Thr	Asp	Val	Ser	Leu	Tyr	Pro	Ser	Gly	Gly	Thr	Gly	
					395					400					405	
	Leu	Glu	Thr	Pro	Asp	Arg	Lys	Gly	Lys	Gly	Thr	Thr	Glu	Gly	Lys	
					410					415					420	
	Pro	Ser	Ser	Phe	Phe	Pro	Glu	Asp	Ser	Phe	Ile	Asp	Ser	Gly	Glu	
					425					430					435	
15	Ile	Asp	Val	Gly	Arg	Arg	Ala	Ser	Gln	Lys	Ile	Pro	Pro	Gly	Thr	
					440					445					450	
	Phe	Trp	Arg	Ser	Gln	Val	Phe	Ile	Asp	His	Pro	Val	His	Leu	Lys	
					455					460					465	
20	Phe	Asn	Val	Ser	Leu	Gly	Lys	Ala	Ala	Leu	Val	Gly	Ile	Tyr	Gly	
					470					475					480	
	Arg	Lys	Gly	Leu	Pro	Pro	Ser	His	Thr	Gln	Phe	Asp	Phe	Val	Glu	
					485					490					495	
	Leu	Leu	Asp	Gly	Arg	Arg	Leu	Leu	Thr	Gln	Glu	Ala	Arg	Ser	Leu	
					500					505					510	
25	Glu	Gly	Thr	Pro	Arg	Gln	Ser	Arg	Gly	Thr	Val	Pro	Pro	Ser	Ser	
					515					520					525	
	His	Glu	Thr	Gly	Phe	Ile	Gln	Tyr	Leu	Asp	Ser	Gly	Ile	Trp	His	
					530					535					540	
30	Leu	Ala	Phe	Tyr	Asn	Asp	Gly	Lys	Glu	Ser	Glu	Val	Val	Ser	Phe	
					545					550					555	
	Leu	Thr	Thr	Ala	Ile	Ala	Leu	Pro	Pro	Arg	Leu	Lys	Glu	Met	Lys	
					560					565					570	
	Ser	Gln	Glu	Ser	Ala	Ala	Gly	Ser	Lys	Leu	Val	Leu	Arg	Cys	Glu	
					575					580					585	
35	Thr	Ser	Ser	Glu	Tyr	Ser	Ser	Leu	Arg	Phe	Lys	Trp	Phe	Lys	Asn	
					590					595					600	
	Gly	Asn	Glu	Leu	Asn	Arg	Lys	Asn	Lys	Pro	Gln	Asn	Ile	Lys	Ile	
					605					610					615	

Gln Lys Lys Pro Gly Lys Ser Glu Leu Arg Ile Asn Lys Ala Ser
 620 625 630
 Leu Ala Asp Ser Gly Glu Tyr Met Cys Lys Val Ile Ser Lys Leu
 635 640 645
 5 Gly Asn Asp Ser Ala Ser Ala Asn Ile Thr Ile Val Glu Ser Asn
 650 655 660
 Glu Ile Ile Thr Gly Met Pro Ala Ser Thr Glu Gly Ala Tyr Val
 665 670 675
 10 Ser Ser Glu Ser Pro Ile Arg Ile Ser Val Ser Thr Glu Gly Ala
 680 685 690
 Asn Thr Ser Ser Ser Thr Ser Thr Ser Thr Thr Gly Thr Ser His
 695 700 705
 Leu Val Lys Cys Ala Glu Lys Glu Lys Thr Phe Cys Val Asn Gly
 710 715 720
 15 Gly Glu Cys Phe Met Val Lys Asp Leu Ser Asn Pro Ser Arg Tyr
 725 730 735
 Leu Cys Lys Cys Pro Asn Glu Phe Thr Gly Asp Arg Cys Gln Asn
 740 745 750
 20 Tyr Val Met Ala Ser Phe Tyr Ser Thr Ser Thr Pro Phe Leu Ser
 755 760 765
 Leu Pro Glu
 768

<210> 12
 <211> 3111
 25 <212> DNA
 <213> Homo sapiens

<400> 12
 gggtagcatg ggtcggtag cgcgtttccc gcctgagcgc aactagcggc 50
 gggtagctggg cacctccaga aaagatcccg caccatcctc caggatccaa 100
 30 tggccttgga gagagggctg cagggccac ggacattgct gactcttcag 150
 aacgtgctga catggagcca ggtagactga aattatcatg tgtccaaatt 200
 aaaattgcat acttcaagga ttatttgaag gactattctt agaccctttt 250
 aagaagattt aaagaaaaac cactcggccc tgagtgcggc gaggaccctg 300
 tttgtggatg tggaggagcg cgggcccggag gcc atg gac gtg 342
 35 Met Asp Val
 1
 aag gag agg aag cct tac cgc tcg ctg acc cgg cgc cgc 381

	Lys	Glu	Arg	Lys	Pro	Tyr	Arg	Ser	Leu	Thr	Arg	Arg	Arg	
	5						10					15		
	gac	gcc	gag	cgc	cgc	tac	acc	agc	tcg	tcc	gcg	gac	agc	420
5	Asp	Ala	Glu	Arg	Arg	Tyr	Thr	Ser	Ser	Ser	Ala	Asp	Ser	
			20					25						
	gag	gag	ggc	aaa	gcc	ccg	cag	aaa	tcg	tac	agc	tcc	agc	459
	Glu	Glu	Gly	Lys	Ala	Pro	Gln	Lys	Ser	Tyr	Ser	Ser	Ser	
	30				35				40					
	gag	acc	ctg	aag	gcc	tac	gac	cag	gac	gcc	cgc	cta	gcc	498
10	Glu	Thr	Leu	Lys	Ala	Tyr	Asp	Gln	Asp	Ala	Arg	Leu	Ala	
		45					50					55		
	tat	ggc	agc	cgc	gtc	aag	gac	att	gtg	ccg	cag	gag	gcc	537
	Tyr	Gly	Ser	Arg	Val	Lys	Asp	Ile	Val	Pro	Gln	Glu	Ala	
				60					65					
15	gag	gaa	ttc	tgc	cgc	aca	ggt	gcc	aac	ttc	acc	ctg	cgg	576
	Glu	Glu	Phe	Cys	Arg	Thr	Gly	Ala	Asn	Phe	Thr	Leu	Arg	
	70					75					80			
	gag	ctg	ggg	ctg	gaa	gaa	gta	acg	ccc	cct	cac	ggg	acc	615
20	Glu	Leu	Gly	Leu	Glu	Glu	Val	Thr	Pro	Pro	His	Gly	Thr	
			85					90						
	ctg	tac	cgg	aca	gac	att	ggc	ctc	ccc	cac	tgc	ggc	tac	654
	Leu	Tyr	Arg	Thr	Asp	Ile	Gly	Leu	Pro	His	Cys	Gly	Tyr	
	95				100				105					
	tcc	atg	ggg	gct	ggc	tct	gat	gcc	gac	atg	gag	gct	gac	693
25	Ser	Met	Gly	Ala	Gly	Ser	Asp	Ala	Asp	Met	Glu	Ala	Asp	
		110					115				120			
	acg	gtg	ctg	tcc	cct	gag	cac	ccc	gtg	cgt	ctg	tgg	ggc	732
	Thr	Val	Leu	Ser	Pro	Glu	His	Pro	Val	Arg	Leu	Trp	Gly	
				125				130						
30	cgg	agc	aca	cgg	tca	ggg	cgc	agc	tcc	tgc	ctg	tcc	agc	771
	Arg	Ser	Thr	Arg	Ser	Gly	Arg	Ser	Ser	Cys	Leu	Ser	Ser	
	135					140					145			
	cgg	gcc	aat	tcc	aat	ctc	aca	ctc	acc	gac	acc	gag	cat	810
35	Arg	Ala	Asn	Ser	Asn	Leu	Thr	Leu	Thr	Asp	Thr	Glu	His	
			150				155							
	gaa	aac	act	gag	act	gat	cat	ccg	ggc	ggc	ctg	cag	aac	849
	Glu	Asn	Thr	Glu	Thr	Asp	His	Pro	Gly	Gly	Leu	Gln	Asn	
	160				165				170					
	cac	gcg	cgg	ctc	cgg	acg	ccg	ccg	ccg	ccg	ctc	tcg	cac	888
40	His	Ala	Arg	Leu	Arg	Thr	Pro	Pro	Pro	Pro	Leu	Ser	His	
		175					180				185			
	gcc	cac	acc	ccc	aac	cag	cac	cac	gcg	gcc	tcc	att	aac	927

	Ala	His	Thr	Pro	Asn	Gln	His	His	Ala	Ala	Ser	Ile	Asn	
					190					195				
	tcc	ctg	aac	cgg	ggc	aac	ttc	acg	ccg	agg	agc	aac	ccc	966
	Ser	Leu	Asn	Arg	Gly	Asn	Phe	Thr	Pro	Arg	Ser	Asn	Pro	
5		200				205						210		
	agc	ccg	gcc	ccc	acg	gac	cac	tcg	ctc	tcc	gga	gag	ccc	1005
	Ser	Pro	Ala	Pro	Thr	Asp	His	Ser	Leu	Ser	Gly	Glu	Pro	
				215					220					
	cct	gcc	ggc	ggc	gcc	cag	gag	cct	gcc	cac	gcc	cag	gag	1044
10	Pro	Ala	Gly	Gly	Ala	Gln	Glu	Pro	Ala	His	Ala	Gln	Glu	
	225					230					235			
	aac	tgg	ctg	ctc	aac	agc	aac	atc	ccc	ctg	gag	acc	aga	1083
	Asn	Trp	Leu	Leu	Asn	Ser	Asn	Ile	Pro	Leu	Glu	Thr	Arg	
			240					245					250	
15	aac	cta	ggc	aag	cag	cca	ttc	cta	ggg	aca	ttg	cag	gac	1122
	Asn	Leu	Gly	Lys	Gln	Pro	Phe	Leu	Gly	Thr	Leu	Gln	Asp	
					255					260				
	aac	ctc	att	gag	atg	gac	att	ctc	ggc	gcc	tcc	cgc	cat	1161
	Asn	Leu	Ile	Glu	Met	Asp	Ile	Leu	Gly	Ala	Ser	Arg	His	
20		265					270					275		
	gat	ggg	gct	tac	agt	gac	ggg	cac	ttc	ctc	ttc	aag	cct	1200
	Asp	Gly	Ala	Tyr	Ser	Asp	Gly	His	Phe	Leu	Phe	Lys	Pro	
				280				285						
	gga	ggc	acc	tcc	ccg	ctc	ttc	tgc	acc	aca	tca	cca	ggg	1239
25	Gly	Gly	Thr	Ser	Pro	Leu	Phe	Cys	Thr	Thr	Ser	Pro	Gly	
	290					295					300			
	tac	cca	ctg	acg	tcc	agc	aca	gtg	tac	tct	cct	ccg	ccc	1278
	Tyr	Pro	Leu	Thr	Ser	Ser	Thr	Val	Tyr	Ser	Pro	Pro	Pro	
				305				310					315	
30	cga	ccc	ctg	ccc	cgc	agc	acc	ttc	gcc	cgg	ccg	gcc	ttt	1317
	Arg	Pro	Leu	Pro	Arg	Ser	Thr	Phe	Ala	Arg	Pro	Ala	Phe	
					320					325				
	aac	ctc	aag	aag	ccc	tcc	aag	tac	tgt	aac	tgg	aag	tgc	1356
	Asn	Leu	Lys	Lys	Pro	Ser	Lys	Tyr	Cys	Asn	Trp	Lys	Cys	
35		330					335					340		
	gca	gcc	ctg	agc	gcc	atc	gtc	atc	tca	gcc	act	ctg	gtc	1395
	Ala	Ala	Leu	Ser	Ala	Ile	Val	Ile	Ser	Ala	Thr	Leu	Val	
				345				350						
	atc	ctg	ctg	gca	tac	ttt	gtg	gcc	atg	cac	ctg	ttt	ggc	1434
40	Ile	Leu	Leu	Ala	Tyr	Phe	Val	Ala	Met	His	Leu	Phe	Gly	
	355					360					365			
	cta	aac	tgg	cac	ctg	cag	ccg	atg	gag	ggg	cag	atg	tat	1473

	Leu	Asn	Trp	His	Leu	Gln	Pro	Met	Glu	Gly	Gln	Met	Tyr	
			370					375					380	
	gag	atc	acg	gag	gac	aca	gcc	agc	agt	tgg	cct	gtg	cca	1512
	Glu	Ile	Thr	Glu	Asp	Thr	Ala	Ser	Ser	Trp	Pro	Val	Pro	
5					385					390				
	acc	gac	gtc	tcc	cta	tac	ccc	tca	ggg	ggc	act	ggc	tta	1551
	Thr	Asp	Val	Ser	Leu	Tyr	Pro	Ser	Gly	Gly	Thr	Gly	Leu	
		395					400					405		
	gag	acc	cct	gac	agg	aaa	ggc	aaa	gga	acc	aca	gaa	gga	1590
10	Glu	Thr	Pro	Asp	Arg	Lys	Gly	Lys	Gly	Thr	Thr	Glu	Gly	
				410				415						
	aag	ccc	agt	agt	ttc	ttt	cca	gag	gac	agt	ttc	ata	gat	1629
	Lys	Pro	Ser	Ser	Phe	Phe	Pro	Glu	Asp	Ser	Phe	Ile	Asp	
	420					425					430			
	tct	gga	gaa	att	gat	gtg	gga	agg	cga	gct	tcc	cag	aag	1668
15	Ser	Gly	Glu	Ile	Asp	Val	Gly	Arg	Arg	Ala	Ser	Gln	Lys	
			435					440					445	
	att	cct	cct	ggc	act	ttc	tgg	aga	tct	caa	gtg	ttc	ata	1707
	Ile	Pro	Pro	Gly	Thr	Phe	Trp	Arg	Ser	Gln	Val	Phe	Ile	
20					450					455				
	gac	cat	cct	gtg	cat	ctg	aaa	ttc	aat	gtg	tct	ctg	gga	1746
	Asp	His	Pro	Val	His	Leu	Lys	Phe	Asn	Val	Ser	Leu	Gly	
		460					465					470		
	aag	gca	gcc	ctg	gtt	ggc	att	tat	ggc	aga	aaa	ggc	ctc	1785
25	Lys	Ala	Ala	Leu	Val	Gly	Ile	Tyr	Gly	Arg	Lys	Gly	Leu	
				475					480					
	cct	cct	tca	cat	aca	cag	ttt	gac	ttt	gtg	gag	ctg	ctg	1824
	Pro	Pro	Ser	His	Thr	Gln	Phe	Asp	Phe	Val	Glu	Leu	Leu	
	485					490					495			
	gat	ggc	agg	agg	ctc	cta	acc	cag	gag	gcg	cgg	agc	cta	1863
30	Asp	Gly	Arg	Arg	Leu	Leu	Thr	Gln	Glu	Ala	Arg	Ser	Leu	
			500					505				510		
	gag	ggg	acc	ccg	cgc	cag	tct	cgg	gga	act	gtg	ccc	ccc	1902
	Glu	Gly	Thr	Pro	Arg	Gln	Ser	Arg	Gly	Thr	Val	Pro	Pro	
35					515					520				
	tcc	agc	cat	gag	aca	ggc	ttc	atc	cag	tat	ttg	gat	tca	1941
	Ser	Ser	His	Glu	Thr	Gly	Phe	Ile	Gln	Tyr	Leu	Asp	Ser	
		525					530					535		
	gga	atc	tgg	cac	ttg	gct	ttt	tac	aat	gac	gga	aag	gag	1980
40	Gly	Ile	Trp	His	Leu	Ala	Phe	Tyr	Asn	Asp	Gly	Lys	Glu	
				540				545						
	tca	gaa	gtg	gtt	tcc	ttt	ctc	acc	act	gcc	att	gcc	ttg	2019

	Ser	Glu	Val	Val	Ser	Phe	Leu	Thr	Thr	Ala	Ile	Ala	Leu	
	550					555					560			
	cct	ccc	cga	ttg	aaa	gag	atg	aaa	agc	cag	gaa	tcg	gct	2058
	Pro	Pro	Arg	Leu	Lys	Glu	Met	Lys	Ser	Gln	Glu	Ser	Ala	
5			565					570					575	
	gca	ggg	tcc	aaa	cta	gtc	ctt	cgg	tgt	gaa	acc	agt	tct	2097
	Ala	Gly	Ser	Lys	Leu	Val	Leu	Arg	Cys	Glu	Thr	Ser	Ser	
					580					585				
	gaa	tac	tcc	tct	ctc	aga	ttc	aag	tgg	ttc	aag	aat	ggg	2136
10	Glu	Tyr	Ser	Ser	Leu	Arg	Phe	Lys	Trp	Phe	Lys	Asn	Gly	
		590					595					600		
	aat	gaa	ttg	aat	cga	aaa	aac	aaa	cca	caa	aat	atc	aag	2175
	Asn	Glu	Leu	Asn	Arg	Lys	Asn	Lys	Pro	Gln	Asn	Ile	Lys	
				605					610					
15	ata	caa	aaa	aag	cca	ggg	aag	tca	gaa	ctt	cgc	att	aac	2214
	Ile	Gln	Lys	Lys	Pro	Gly	Lys	Ser	Glu	Leu	Arg	Ile	Asn	
	615					620					625			
	aaa	gca	tca	ctg	gct	gat	tct	gga	gag	tat	atg	tgc	aaa	2253
	Lys	Ala	Ser	Leu	Ala	Asp	Ser	Gly	Glu	Tyr	Met	Cys	Lys	
20			630					635					640	
	gtg	atc	agc	aaa	tta	gga	aat	gac	agt	gcc	tct	gcc	aat	2292
	Val	Ile	Ser	Lys	Leu	Gly	Asn	Asp	Ser	Ala	Ser	Ala	Asn	
					645					650				
	atc	acc	atc	gtg	gaa	tca	aac	gag	atc	atc	act	ggg	atg	2331
25	Ile	Thr	Ile	Val	Glu	Ser	Asn	Glu	Ile	Ile	Thr	Gly	Met	
		655					660					665		
	cca	gcc	tca	act	gaa	gga	gca	tat	gtg	tct	tca	gag	tct	2370
	Pro	Ala	Ser	Thr	Glu	Gly	Ala	Tyr	Val	Ser	Ser	Glu	Ser	
				670				675						
30	ccc	att	aga	ata	tca	gta	tcc	aca	gaa	gga	gca	aat	act	2409
	Pro	Ile	Arg	Ile	Ser	Val	Ser	Thr	Glu	Gly	Ala	Asn	Thr	
	680					685					690			
	tct	tca	tct	aca	tct	aca	tcc	acc	act	ggg	aca	agc	cat	2448
	Ser	Ser	Ser	Thr	Ser	Thr	Ser	Thr	Thr	Gly	Thr	Ser	His	
35			695					700					705	
	ctt	gta	aaa	tgt	gcg	gag	aag	gag	aaa	act	ttc	tgt	gtg	2487
	Leu	Val	Lys	Cys	Ala	Glu	Lys	Glu	Lys	Thr	Phe	Cys	Val	
					710					715				
	aat	gga	ggg	gag	tgc	ttc	atg	gtg	aaa	gac	ctt	tca	aac	2526
40	Asn	Gly	Gly	Glu	Cys	Phe	Met	Val	Lys	Asp	Leu	Ser	Asn	
		720					725					730		
	ccc	tcg	aga	tac	ttg	tgc	aag	tgc	cca	aat	gag	ttt	act	2565

Pro Ser Arg Tyr Leu Cys Lys Cys Pro Asn Glu Phe Thr
 735 740

ggt gat cgc tgc caa aac tac gta atg gcc agc ttc tac 2604
 Gly Asp Arg Cys Gln Asn Tyr Val Met Ala Ser Phe Tyr
 5 745 750 755

agt acg tcc act ccc ttt ctg tct ctg cct gaa tag 2640
 Ser Thr Ser Thr Pro Phe Leu Ser Leu Pro Glu
 760 765 768

gagcatgctc agttggtgct gctttcttgt tgctgcatct cccctcagat 2690

10 tccacctaga gctagatgtg tcttaccaga tctaataattg actgcctctg 2740

cctgtcgcat gagaacatta acaaaagcaa ttgtattact tectctgttc 2790

gcgactagtt ggctctgaga tactaatagg tgtgtgaggc tccggatgtt 2840

tctggaattg atattgaatg atgtgataca aattgatagt caatatcaag 2890

cagtgaata tgataataaa ggcatttcaa agtctcactt ttattgataa 2940

15 aataaaaaatc attctactga acagtccatc ttcttttatac aatgaccaca 2990

tcctgaaaag ggtgttgcta agctgtaacc gatatgcact tgaaatgatg 3040

gtaagttaat ttgattcag aatgtgttat ttgtcacaaa taaacataat 3090

aaaaggaaaa aaaaaaaaaa a 3111

<210> 13
 20 <211> 1872
 <212> DNA
 <213> Homo sapiens

<400> 13

gaattcggga cagcctctcc tgccgcgct gctgctgccg ccgccgccac 50

25 cgccggctgg tctccttct gcttttactt ctctgcatg acagttgttt 100

tcttcatctg agcagacacc agcttcagat gctcgagggtg agaaacatgc 150

ctttcagttt gggctactgg ttacttaat taatcagccg gcagctccgt 200

cgatctattt tcgtccctgt cctcttgacg agcccgggat ggtttgaggt 250

agcatttaaa agaactagaa aagtggccca gaaacagcag cttaaagaat 300

30 tattacgata tactttgatt ttgtagttgc taggagcttt tcttcccccc 350

ttgcatcttt ctgaactctt cttgatttta ataatggcct tggacttgga 400

cgatttatcg atttccccct gtaagatgct gtatcatttg gttggggggg 450

cctctgcgtg gtaatggacc gtgagagcgg ccaggccttc ttctggaggt 500

```

gagccg  atg gag att tat tcc cca gac atg tct gag gtc 539
          Met Glu Ile Tyr Ser Pro Asp Met Ser Glu Val
            1             5             10

gcc gcc gag agg tcc tcc agc ccc tcc act cag ctg agt 578
5  Ala Ala Glu Arg Ser Ser Ser Pro Ser Thr Gln Leu Ser
            15             20

gca gac cca tct ctt gat ggg ctt ccg gca gca gaa gac 617
Ala Asp Pro Ser Leu Asp Gly Leu Pro Ala Ala Glu Asp
  25             30             35

10  atg cca gag ccc cag act gaa gat ggg aga acc cct gga 656
    Met Pro Glu Pro Gln Thr Glu Asp Gly Arg Thr Pro Gly
            40             45             50

ctc gtg ggc ctg gcc gtg ccc tgc tgt gcg tgc cta gaa 695
15  Leu Val Gly Leu Ala Val Pro Cys Cys Ala Cys Leu Glu
            55             60

gct gag cgc ctg aga ggt tgc ctc aac tca gag aaa atc 734
    Ala Glu Arg Leu Arg Gly Cys Leu Asn Ser Glu Lys Ile
            65             70             75

20  tgc att gtc ccc atc ctg gct tgc ctg gtc agc ctc tgc 773
    Cys Ile Val Pro Ile Leu Ala Cys Leu Val Ser Leu Cys
            80             85

ctc tgc atc gcc ggc ctc aag tgg gta ttt gtg gac aag 812
    Leu Cys Ile Ala Gly Leu Lys Trp Val Phe Val Asp Lys
            90             95             100

25  atc ttt gaa tat gac tct cct act cac ctt gac cct ggg 851
    Ile Phe Glu Tyr Asp Ser Pro Thr His Leu Asp Pro Gly
            105             110             115

ggg tta ggc cag gac cct att att tct ctg gac gca act 890
30  Gly Leu Gly Gln Asp Pro Ile Ile Ser Leu Asp Ala Thr
            120             125

gct gcc tca gct gtg tgg gtg tcg tct gag gca tac act 929
    Ala Ala Ser Ala Val Trp Val Ser Ser Glu Ala Tyr Thr
            130             135             140

35  tca cct gtc tct agg gct caa tct gaa agt gag gtt caa 968
    Ser Pro Val Ser Arg Ala Gln Ser Glu Ser Glu Val Gln
            145             150

gtt aca gtg caa ggt gac aag gct gtt gtc tcc ttt gaa 1007
    Val Thr Val Gln Gly Asp Lys Ala Val Val Ser Phe Glu
            155             160             165

40  cca tca gcg gca ccg aca ccg aag aat cgt att ttt gcc 1046
    Pro Ser Ala Ala Pro Thr Pro Lys Asn Arg Ile Phe Ala
            170             175             180

```


ttt tct ttc ttg ccg tcc act gcg cca tcc ttc cct tca 1085
 Phe Ser Phe Leu Pro Ser Thr Ala Pro Ser Phe Pro Ser
 185 190

5 ccc acc cgg aac cct gag gtg aga acg ccc aag tca gca 1124
 Pro Thr Arg Asn Pro Glu Val Arg Thr Pro Lys Ser Ala
 195 200 205

act cag cca caa aca aca gaa act aat ctc caa act gct 1163
 Thr Gln Pro Gln Thr Thr Glu Thr Asn Leu Gln Thr Ala
 210 215

10 cct aaa ctt tct aca tct aca tcc acc act ggg aca agc 1202
 Pro Lys Leu Ser Thr Ser Thr Ser Thr Thr Gly Thr Ser
 220 225 230

cat ctt gta aaa tgt gcg gag aag gag aaa act ttc tgt 1241
 15 His Leu Val Lys Cys Ala Glu Lys Glu Lys Thr Phe Cys
 235 240 245

gtg aat gga ggg gag tgc ttc atg gtg aaa gac ctt tca 1280
 Val Asn Gly Gly Glu Cys Phe Met Val Lys Asp Leu Ser
 250 255

aac ccc tcg aga tac ttg tgc aag tgc cca aat gag ttt 1319
 20 Asn Pro Ser Arg Tyr Leu Cys Lys Cys Pro Asn Glu Phe
 260 265 270

act ggt gat cgc tgc caa aac tac gta atg gcc agc ttc 1358
 Thr Gly Asp Arg Cys Gln Asn Tyr Val Met Ala Ser Phe
 275 280

25 tac agt acg tcc act ccc ttt ctg tct ctg cct gaa taggag 1400
 Tyr Ser Thr Ser Thr Pro Phe Leu Ser Leu Pro Glu
 285 290 295 296

catgctcagt tgggtgctgct ttcttgttgc tgcattctccc ctcagattcc 1450
 acctagagct agatgtgtct taccagatct aatattgact gcctctgcct 1500

30 gtcgcatgag aacattaaca aaagcaattg tattacttcc tctgttcgcg 1550
 actagtggc tctgagatac taataggtgt gtgaggctcc ggatgtttct 1600
 ggaattgata ttgaatgatg tgatacaaat tgatagtcaa tatcaagcag 1650
 tgaaatatga taataaaggc atttcaaagt ctcactttta ttgataaaat 1700
 aaaaatcatt ctactgaaca gtccatcttc tttatacaat gaccacatcc 1750

35 tgaaaagggt gttgctaagc tgtaaccgat atgcacttga aatgatggta 1800
 agttaatttt gattcagaat gtgttatttg tcacaaataa acataataaa 1850
 aggaaaaaaa aaaccogaat tc 1872

<210> 14
 <211> 296
 <212> PRT
 <213> Homo sapiens

5 <400> 14
 Met Glu Ile Tyr Ser Pro Asp Met Ser Glu Val Ala Ala Glu Arg
 1 5 10 15
 Ser Ser Ser Pro Ser Thr Gln Leu Ser Ala Asp Pro Ser Leu Asp
 20 25 30
 10 Gly Leu Pro Ala Ala Glu Asp Met Pro Glu Pro Gln Thr Glu Asp
 35 40 45
 Gly Arg Thr Pro Gly Leu Val Gly Leu Ala Val Pro Cys Cys Ala
 50 55 60
 Cys Leu Glu Ala Glu Arg Leu Arg Gly Cys Leu Asn Ser Glu Lys
 15 65 70 75
 Ile Cys Ile Val Pro Ile Leu Ala Cys Leu Val Ser Leu Cys Leu
 80 85 90
 Cys Ile Ala Gly Leu Lys Trp Val Phe Val Asp Lys Ile Phe Glu
 95 100 105
 20 Tyr Asp Ser Pro Thr His Leu Asp Pro Gly Gly Leu Gly Gln Asp
 110 115 120
 Pro Ile Ile Ser Leu Asp Ala Thr Ala Ala Ser Ala Val Trp Val
 125 130 135
 Ser Ser Glu Ala Tyr Thr Ser Pro Val Ser Arg Ala Gln Ser Glu
 25 140 145 150
 Ser Glu Val Gln Val Thr Val Gln Gly Asp Lys Ala Val Val Ser
 155 160 165
 Phe Glu Pro Ser Ala Ala Pro Thr Pro Lys Asn Arg Ile Phe Ala
 170 175 180
 30 Phe Ser Phe Leu Pro Ser Thr Ala Pro Ser Phe Pro Ser Pro Thr
 185 190 195
 Arg Asn Pro Glu Val Arg Thr Pro Lys Ser Ala Thr Gln Pro Gln
 200 205 210
 Thr Thr Glu Thr Asn Leu Gln Thr Ala Pro Lys Leu Ser Thr Ser
 35 215 220 225
 Thr Ser Thr Thr Gly Thr Ser His Leu Val Lys Cys Ala Glu Lys
 230 235 240
 Glu Lys Thr Phe Cys Val Asn Gly Gly Glu Cys Phe Met Val Lys
 245 250 255

Asp Leu Ser Asn Pro Ser Arg Tyr Leu Cys Lys Cys Pro Asn Glu
260 265 270

Phe Thr Gly Asp Arg Cys Gln Asn Tyr Val Met Ala Ser Phe Tyr
275 280 285

5 Ser Thr Ser Thr Pro Phe Leu Ser Leu Pro Glu
290 295 296